Draft

Initial Calibration and Data Gap Analysis Report

Dual Site Groundwater Operable Unit Remedial Design Montrose Chemical and Del Amo Superfund Sites

Prepared for

Environmental Protection Agency Region IX

75 Hawthorne Street
San Francisco, California

October 2004

CH2MHILL
3 Hutton Centre Drive, Suite 200
Santa Ana, California 92707

Contents

Appendix

Tables

Figures

Acronyms

DNAPL	dense nonaqueous-phase liquid
EPA	United States Environmental Protection Agency
ft/day	feet per day
ft/ft	foot per foot
GHB	general head boundary
GLA	Gage-Lynwood Aquitard
gpm	gallons per minute
HSU	hydrostratigraphic unit
in/yr	inches per year
ISGS	in situ groundwater standards
JGWFS	Joint Groundwater Feasibility Study
K	hydraulic conductivity
Kd	distribution coefficient
Kh	horizontal hydraulic conductivity
Kv	vertical hydraulic conductivity
LBF	Lower Bellflower Aquitard
µg/L	micrograms per liter
MBFB	Middle Bellflower B-Sand
MBFC	Middle Bellflower C-Sand
MBFM	Middle Bellflower Mud
MCL	maximum contaminant level
mL/g	milliliters per gallon
NAPL	nonaqueous-phase liquid
p-CBSA	parachlorobenzene sulfonic acid
ppb	parts per billion
R	retardation factor
RD	remedial design

RI	remedial investigation
RMS	root mean squared
ROD	Record of Decision
Site	Montrose/Del Amo Dual Site
TCE	trichloroethylene
TI	Technical Impracticability
UBF	Upper Bellflower Aquitard
USGS	United States Geological Survey

1. Introduction

This report presents the results of the initial calibration of the Remedial Design (RD) model and data gap analysis performed as part of the Dual Site Groundwater Operable Unit RD for the Montrose Chemical and Del Amo Superfund Sites (Site). The original numerical model of the Site was developed as part of the Joint Groundwater Feasibility Study (JGWFS) to compare remedial alternatives. Although the JGWFS model was based on the results of extensive remedial investigations, there were a number of uncertainties associated with the model input parameters that were caused by the inherent complexity of the hydrogeologic conditions beneath the Site. While these uncertainties are not uncommon when modeling complex physical systems such as the Site, they need to be evaluated, quantified, and reduced (to the extent possible) to increase the accuracy of modeling predictions for the design of the remedial wellfield(s). To achieve this goal, an initial calibration and data gap analysis was performed as part of the RD to quantify the predictive uncertainty of the RD model, and to identify data types that could have the greatest effect in reducing this uncertainty (i.e., identify data gaps). The methodology for this analysis is discussed in Tasks 2.3 and 2.4 of the Work Plan for Model Development (CH2M HILL, 2003).

The groundwater flow model MODFLOW2000 (United States Geological Survey [USGS], 2000), the solute-transport model MT3DMS (Zheng and Wang, 1999), and the particle-tracking code MODPATH (USGS, 1994) were used for the development of the RD model. MODFLOW2000 and MT3DMS are the updated versions of the numerical flow and transport codes that were used for the JGWFS model (i.e., MODFLOW [USGS, 1988] and MT3D96 [S.S. Papadopulos, 1996]). These codes are widely accepted by the regulatory community, and are used extensively by the United States Environmental Protection Agency (EPA) at numerous sites across the country, primarily because these codes are in the public domain, are well-documented, and have been verified against a number of analytical solutions.

The model domain, locations of boundary conditions, and model grid (other than the model-top modification noted in Section 3) of the JGWFS model were used for the purposes of the initial calibration/data gap analysis, because no additional interpretation of hydrostratigraphic data was performed at this stage. Similar to the JGWFS model, the steady state flow condition was assumed for the purposes of this analysis because no additional interpretation of water level trends and fluxes was performed at this stage. The appropriateness of the model geometry and steady state flow conditions for the final RD model will be assessed based on the results of ongoing data acquisition programs. The JGWFS model is discussed in detail in the Final JGWFS for the Montrose and Del Amo Sites, Appendix B (CH2M HILL, 1998).

Initial calibration, evaluation of model uncertainties, and data gap analysis were performed through the use of a systematic calibration process utilizing the automatic calibration software package PEST (Doherty, 2002; Doherty and Johnston, 2003). Results of this analysis were used to develop recommendations for collecting additional data in the field, which would help to reduce the predictive uncertainty of the model with regard to the performance of the remedial wellfield.

All modeling input and output files and a brief documentation of the results generated for this data gap analysis were posted on the File Transfer Protocol (FTP) site for review. It is important to note that the model simulations were performed in this study only for the purposes of initial calibration/data gap analysis, and are not intended as the ultimate design of the remedial wellfield. The objectives of the initial calibration/data gap analysis and the organization of this report are discussed below.

1.1 Objectives of Initial Calibration and Data Gap Analysis

As mentioned above, the overall objective of the initial calibration/data gap analysis is to quantify the predictive uncertainty of the RD model, and to identify data types that could have the greatest effect in reducing this uncertainty. The specific objectives of this work are listed below:

- Create the needed connections and interface, and write needed routines, so that PEST will run with the MODFLOW2000, MT3DMS, and MODPATH modeling codes for the RD model domain and structure.
- Perform an initial "baseline" calibration to a set of observed and measured calibration targets such as water levels, contaminant concentrations, etc.
- Develop a set of calibrated models that imply large differences in remedial results from success to failure.
- Evaluate the predictive uncertainty of the model with regard to the performance of the remedial wellfield, and assess the contribution to this uncertainty from various model parameters.
- Use the results of this evaluation to determine where the uncertainty may be reducible as opposed to irreducible, and which additional field data (e.g., aquifer tests at monitoring wells, additional well installations, etc.) are warranted.
- Provide a modeling tool that is ready to use to recalibrate the model quickly as new data are collected.

1.2 Report Organization

The report consists of the following sections:

Section 1, Introduction – Presents the objectives of this analysis and the report organization.

Section 2, Calibration Methodology – Discusses the calibration methodology using PEST, including initial "baseline" calibration as well as the predictive calibration and uncertainty analysis.

Section 3, Initial Calibration – Discusses the assumed distribution and limits for model parameters, the calibration targets and target weights, and the initial calibration results.

Section 4, Data Gap Analysis – Discusses the predictive uncertainty of the model with regard to the performance of the remedial wellfield, such as (1) the mass and volume reduction rates of the chlorobenzene plume, and (2) the hydraulic containment of the

chlorobenzene plume. The assessment of the model predictive uncertainty was used to identify data types that could have the greatest effect in reducing this uncertainty.

Section 5, Conclusions and Recommendations – Presents the conclusions and recommendations of the data gap analysis regarding additional field data that need to be collected to reduce the uncertainty of modeling predictions.

Section 6, References – Presents a list of bibliographic references used in this report.

2. Calibration Methodology

The initial calibration of the RD model and the data gap analysis were performed using the nonlinear parameter estimation software package PEST (Doherty, 2002; Doherty and Johnston, 2003). The methodology for the initial calibration, which is also referred to in this report as the "baseline calibration," and for the data gap analysis is discussed below.

2.1 Initial Calibration Methodology

As discussed in detail in the Work Plan for the Development of the Groundwater Model for the Remedial Design (CH2M HILL, 2003), the model calibration using PEST is performed by automatic minimization of the objective function, which is the sum of squared residuals of the calibration targets. Calibration targets are observed or estimated parameters such as water levels, contaminant concentrations, etc. (see Section 3.2). Residuals are the differences between model-simulated and observed conditions. In the process of calibration, PEST modifies calibration parameters (e.g., hydraulic conductivity, river conductance, etc.) in accordance with the prescribed parameter distribution and limits (Section 3.1) until the minimum objective function is achieved (i.e., the objective function can no longer be reduced). To achieve a minimum objective function, PEST first runs the model in its original condition and determines the total objective function. PEST then automatically makes a small change to the first parameter, runs the model, and determines the change to the objective function. Next, PEST changes the first parameter back to its original value, makes a small change to the second parameter, and so on. When that process is complete, the change to the objective function caused by each parameter is used to solve for a new set of parameter values. The new parameter values are then used as a starting point to repeat the process. When the objective function can no longer be reduced, the process is complete.

Because groundwater flow and transport models are generally nonunique, changes to certain parameters can be offset by changes to other parameters, resulting in a similarly reduced objective function, and therefore a similar quality of calibration. Hence, a number of models can be developed using PEST, all of which would be reasonably well calibrated and based on equally viable hydrogeologic parameters for a given physical system.

A number of methods (all under the broad name "regularization") are used to obtain a unique calibration solution. PEST uses "Tichonov Regularization," in which a "preferred condition" for each calibration parameter needs to be defined by the regularization equations (see Section 3.1). The preferred condition could be either a "smoothness condition" (i.e., the condition that is based on the assumption that the geologic media is homogeneous and heterogeneity is introduced only when it is absolutely necessary to meet calibration targets), and/or a set of preferred parameter values, from which deviations are tolerable only to the extent that they are supported by the data. A unique calibration solution is obtained as the result of this regularization process. However, this solution is unique only for that particular regularization mechanism. Should another regularization mechanism be used (for example, different parameters governing the smoothness condition,

different preferred parameter values, and/or different set of targets), then another solution to the inverse problem will be obtained, which could be entirely different from the first.

The initial calibration was based only on one set of preferred conditions (i.e., regularization equations) and calibration targets. Consequently, only one solution (i.e., baseline calibration) was obtained as the result of the initial calibration. This baseline calibration solution was used as a basis for predictive calibration, which was performed to assess other potential solutions to the calibration problem and the impact of these different solutions to the predictive outcome (see Section 2.2). The assumed parameter distribution, limits, preferred values, calibration targets, and the calibration results for the initial calibration are discussed in Section 3.

2.2 Predictive Calibration

As discussed in Section 2.1, a number of models can be developed using PEST, all of which would be reasonably well calibrated and based on equally viable hydrogeologic parameters for a given physical system. These calibrated models may differ, however, with regard to predictions pertaining to the performance of the remedial wellfield. There are several methods that could be used to assess the range of model predictions (i.e., assess the predictive uncertainty of the model).

One of the most comprehensive methods discussed in the Work Plan for the Development of the Groundwater Model (CH2M HILL, 2003) involves Monte Carlo analysis, in which each calibration parameter is assigned a random value, and a set of stochastic fields is generated – each based on what is known about the amount of heterogeneity prevailing within an area. The randomized model consisting of these stochastic fields is then sent through the PEST calibration process in order to ensure that its associated objective function is acceptable. This full stochastic analysis would provide the most comprehensive information and most quantitative assessment of the modeling uncertainty. In such an analysis, PEST would run hundreds of models (each requiring about 1,000 individual runs to calibrate) with differing stochastic parameter fields. The trade-offs are that this analysis is the most costly and time-consuming, and most subject to issues with numerical stability.

The Work Plan for the Model Development (CH2M HILL, 2003) stated that depending on the numerical behavior of the flow and transport models, this or other methods may be implemented to assess the predictive uncertainty. During the initial calibration, it was determined that the MT3DMS code was not stable enough to support a full stochastic analysis in the RD model. Therefore, an alternative method of predictive calibration was selected for the analysis of predictive uncertainty. The use of this method greatly reduced the cost and duration of the modeling effort. However, the resulting evaluation of modeling uncertainty is somewhat less quantitative than what could be obtained by a full stochastic analysis, though it is still significantly superior to the knowledge of uncertainty attainable using only the original JGWFS model.

Predictive calibration involves the use of "predictive targets" in the calibration process in addition to the calibration targets. The predictive targets are based on the standards, requirements, and specifications for the remedial actions outlined in Section 13 of the Record of Decision (ROD) (EPA, 1999) and include the rates of volume and mass reduction in the aquifer as well as the plume containment targets (see Sections 4.1 and 4.2). In the

process of predictive calibration, the model is run by PEST in both calibration and predictive modes. PEST is provided with an objective function, which is the sum of squared residuals of both the calibration and predictive targets. The value of this objective function is slightly higher than that for the baseline calibration (i.e., best calibration that PEST achieved during the initial calibration process). For example, if the total objective function at the end of the baseline calibration was 100, PEST might be allowed to reach an objective function as high as 102 while meeting both the calibration and predictive targets. This process results in a model that is, essentially, just as well-calibrated, but possesses model parameters that result in meeting the predictive targets.

Several combined "calibration/prediction" runs were performed in the process of predictive calibration to identify different combinations of model parameters that would cause opposite predictive outcomes (i.e., possible best and worst outcomes) with regard to the performance of the remedial wellfield, while satisfying the calibration constraints. This was achieved by using the same calibration targets and different sets of predictive targets (i.e., one set of predictive targets prescribed success, and another set of targets prescribed failure of the remedial wellfield to meet certain design criteria). The objective of these simulations was to obtain the maximum range of possible predictions of the calibrated model(s) with regard to the performance of the remedial wellfield (i.e., assess predictive uncertainty of the model). For example, a predictive target of partial containment was used to identify a set of viable model parameters resulting in both model calibration and the failure of the remedial wellfield to contain the plume (see Section 4.2). Conversely, a predictive target of complete plume containment was used to determine a set of calibrated parameters that resulted in adequate performance of the remedial wellfield with regard to containment. If the best- and worst-case simulations were similar in predictive outcome (as was determined for the plume containment in the Gage Aquifer, where plume containment was achieved under different sets of predictive targets), the predictive uncertainty of the model was considered to be acceptable, and the existing data were considered sufficient for this particular modeling prediction. However, if the simulated results varied significantly between the possible best and worst cases for meeting certain remedial design criteria, then the most critical parameters with regard to the model predictions of this criteria were identified and recommended for further assessment to the extent possible.

The identification of the most critical parameters affecting the performance of the remedial wellfield was performed based on several qualitative and quantitative analyses, which included the qualitative comparison of calibrated model parameter distributions, evaluation of the sensitivity of the predictive targets to each model parameter, and evaluation of the contribution of each parameter to the predictive uncertainty of the model.

There are generally two primary sources of parameter uncertainty in the calibrated model. First, there are often an infinite number of ways that parameters could be varied while still maintaining the model in a perfectly calibrated state. This results from the inherent nonuniqueness of the inversion process, where a large number of parameters are calibrated against calibration targets. The second source of uncertainty comes from measurement noise (i.e., potential measurement errors in calibration targets). The estimates of parameter contribution to the model predictive uncertainty performed as part of this data gap analysis considered both of these sources of parameter uncertainty. A methodology for these estimates is described in *The Role of the Calibration Process in Reducing Model Predictive Error* (Moore and Doherty, 2004) presented in Appendix A. This methodology takes into account

potential measurement noise and variability of model parameters at a level of detail that is sufficient to influence model predictions, but is too fine to be captured by the calibration process. The contribution of each parameter to predictive uncertainty is estimated as predictive error variance based on (1) the sensitivity of the prediction to changes in the parameter estimated using PEST during calibration, and (2) the probability distribution (i.e., assumed statistical distribution) of the parameter (see Section 3.1.1). This approach is more comprehensive then simply ranking parameter sensitivities, because a very sensitive parameter may have a small standard deviation, and thus result in little reduction of uncertainty if it becomes more accurately known (i.e., gets measured in the field). Combining the sensitivity with the statistical distribution is essentially a normalizing process, resulting in variances that are directly comparable, and additive. The total predictive error variance is the sum of the contributions from each parameter.

Parameters that had the highest contribution to the model uncertainty with regard to a particular prediction were identified in terms of (1) type (e.g., hydraulic conductivity, boundary conditions, etc.), (2) aquifer, and (3) location (to the extent possible). This information was then used to provide recommendations for additional data collection programs (see Section 5).

3. Initial Calibration

The initial "baseline" calibration of the Site model represents one of the multiple calibration solutions, which provides a reasonable match between the observed, measured, and estimated parameters (i.e., calibration targets) and simulated conditions. This model was developed on the basis of the JGWFS model (i.e., using MODFLOW2000 and MT3DMS numerical codes, grid and layering of the JGWFS model, steady-state flow and transient transport calibration, and general head boundaries). The transient transport run was performed for a period of 57 years, from 1945 (the assumed time of contaminant release to groundwater) through 2002 (the latest period for which the concentration data were available at the time of initial calibration). The initial concentrations of chlorobenzene and parachlorobenzene sulfonic acid (p-CBSA) for the calibration run were assumed to be equal to zero outside of the source area located at the Montrose property. The simulated locations of the source terms were similar to those of the JGWFS model.

In the JGWFS model, model layers could convert between confined and unconfined conditions. Nonlinear behavior introduced by wetting and drying model cells, a known feature of the JGWFS model, commonly results in convergence problems for PEST. To increase solution stability and reduce model run times, the top of the RD model was modified to reflect the approximate location of the water table, rather than the land surface as in the JGWFS model. As part of this process, all layers of the model were converted to confined conditions. This modification is considered acceptable, because the water table beneath the site occurs within the aquitard material (i.e., the Upper Bellflower Aquitard [UBF] or Middle Bellflower B-Sand [MBFB]) composed mostly of interbedded silty sands and silts. Silty interbeds overlaying the water table could potentially provide localized confinement. In addition, because the focus of the remedial actions is the deeper confined aquifers, such as the Middle Bellflower C-Sand (MBFC) and the Gage Aquifer, the conversion of the water table aquifer to confined conditions is not expected to have an effect on the simulations of the remedial wellfield.

The grid, layering, and position of sources in the JGWFS model are described in detail in the JGWFS for the Del Amo and Montrose Site (CH2M HILL, 1998). Table 3-1 shows the correlation between the hydrostratigraphic units beneath the site and model layers.

In the process of the baseline calibration, PEST was given (1) parameter distribution (e.g., fixed or variable, etc.), limits, and preferred values for model parameters, and (2) a set of calibration targets and weights (see Sections 3.1 and 3.2). As mentioned above, the numerical model files for the baseline calibration were provided at the CH2M HILL File Transfer Protocol (FTP) site.

TABLE 3-1
Hydrostratigraphic Units and Model Layers

Hydrostratigraphic Units (HSUs) - #			Model Layers
Del Amo Nomenclature		**Montrose Nomenclature**	**Model Layers**
Upper Bellflower Aquitard (UBF) – 1		Upper Bellflower Aquitard (UBA)	1
Middle Bellflower Aquitard (MBF)	MBFB – 2		2
	MBFM – 3		3
	MBFC – 4	Bellflower Sand	4 and 5
Lower Bellflower Aquitard (LBF) – 5		Lower Bellflower Aquitard (LBF)	6, 7, 8
Gage Aquifer – 6		Gage Aquifer	9
Gage-Lynwood Aquitard (GLA) – 7		Gage-Lynwood Aquitard (GLA)	10, 11, 12
Lynwood Aquifer – 8		Lynwood Aquifer	13

3.1 Parameter Distribution and Limits

As described below, the values of some model parameters were fixed (i.e., were not allowed to change in the calibration process), while others, referred to as calibration parameters, were allowed to vary to achieve the best match between the simulated conditions and calibration targets. Fixed parameters were assigned values based on field tests and estimates performed during the remedial investigations (RI) and JGWFS (e.g., horizontal hydraulic conductivity values were fixed at the locations of aquifer pump tests) and/or literature and professional judgement where field data were not available. Calibration parameters were estimated using PEST. These estimates were performed in accordance with the parameter distribution and limits assigned to each calibration parameter. For some parameters, which were assumed to be constant within a hydrostratigraphic unit (HSU) and/or model layer, a single value was estimated per model layer or HSU. Other parameters were estimated using pilot points. As discussed in the Work Plan for Model Development (CH2M HILL, 2003), pilot points are discrete locations where PEST estimates values of the particular calibration parameter needed to match calibration targets at this location. The values of calibration parameters at all model cells were then interpolated based on the values at the pilot points.

The parameter distribution, limits, and preferred values assumed for the baseline calibration are presented below. Note that the assumptions regarding the model parameters were made for the purposes of this data gap analysis based on the existing data collected during RI and interpretations made for the JGWFS model. These assumptions will be re-evaluated for the RD model based on the results of the ongoing data acquisition programs.

3.1.1 Parameter Distribution and Variability

The assumptions regarding the distribution and variability of calibration parameters and parameters, for which the values were fixed, for the baseline calibration are described below:

Distribution of Calibration Parameters

The assumed spatial distribution for each parameter is presented in Table 3-2 and described below.

- Horizontal hydraulic conductivity (Kh), Kh and vertical hydraulic conductivity (Kv) ratio (Kh:Kv) of upper units, recharge, and Dominguez Channel conductance were assumed to be spatially variable (i.e., were allowed to vary within hydrostratigraphic units [HSUs]).

- Transport parameters such as porosity, dispersivity, distribution coefficient (Kd), retardation factor (R), and source concentrations were assumed to be constant within HSUs, but were allowed to vary in different HSUs.

- General head boundary (GHB) conductance was calculated from the calibrated hydraulic conductivity values (see Section 3.1.2).

As discussed in Section 2.1, the spatial continuity targets were added to the objective function to achieve a homogeneous distribution of parameters unless suggested otherwise by field data used as calibration targets.

Variability of Calibration Parameters

As described in Moore and Doherty (2004), estimates of the parameter contribution to predictive uncertainty of the model require that the probability distribution (i.e., standard deviation) be supplied for each model parameter (Appendix A). To simplify the mathematical procedure for these estimates, statistical independence of spatial parameter variability was assumed for the purposes of this analysis. This assumption is reasonable with regard to the parameters that are not allowed to vary spatially within HSUs (i.e., constant in each HSU), as vertical correlation is generally very weak in geologic materials. For parameters represented by pilot points (i.e., parameters allowed to vary within HSUs) spatial correlation is possible. Nevertheless, the assumption of statistical independence was adopted for these parameters because the use of pilot points as a parameterization device results in a "region of influence" around each of these points, which is of the same order as pilot point separation distances. Thus, spatial correlation is already "built into" the parameterization scheme. The assumed transformation status and standard deviation for each model parameter is presented in Table 3-2.

TABLE 3-2
Assumed Parameter Distribution, Transformation Status, and Standard Deviation

Parameter	Spatial Distribution	Transformation Status	Standard Deviation (Transformed Parameter)
Porosity	Constant in each HSU	None	0.05
Longitudinal dispersivity	Constant in each HSU	Log	0.5
Transverse dispersivity	Constant in each HSU	Log	0.2
Vertical Dispersivity	Constant in each HSU	Log	0.2
Chlorobenzene source concentration	Constant within source area in each HSU	Log	1.0
Chlorobenzene Kd	Constant within each layer	Log	1.0
p-CBSA source concentration	Constant within source area in each HSU	Log	1.5
Riverbed conductance	Four pilot points along river length	Log	0.05
Elevation of general head boundary	Six pilot points along boundary in each aquifer	None	1.5 (ft)
Recharge	Pilot points	None	0.02 (ft/yr)
Hydraulic conductivity	Pilot points within each HSU	Log	0.5

Fixed Parameters

- Values of Kh were fixed (i.e., were not allowed to change in the calibration process) at aquifer test locations.
- Values of Kh:Kv were fixed for lower HSUs (the Middle Bellflower Mud [MBFM] through Lynwood Aquifer).
- Values of GHB heads were estimated based on the extrapolation of available water levels and fixed.
- Source timing was fixed.

3.1.2 Parameter Limits and Preferred Values

Parameter limits and preferred values for flow and transport parameters are discussed below.

Parameter Limits for Flow Parameters

- Kh was allowed to range within plus/minus two standard deviations (log-transformed) of the geometric mean of available aquifer test data, or within plus/minus 1.5 orders of magnitude of the JGWFS model value in the HSUs for which aquifer test data are not

available. Some HSUs without aquifer test data were assigned multiple values in the JGWFS model, in which case a harmonic mean was calculated as the central value.

- Kh:Kv was allowed to range from 10:1 to 3,162:1 in the UBF (+0.5/-2 orders of magnitude from a central value of 1,000, which is not unusual for heterogeneous formations composed of interbedded sands, silts, and clays such as UBF), from 10:1 to 1,000:1 in the MBFB, and was fixed at 10:1 in lower units (i.e., in the MBFM through Lynwood Aquifer).
- Recharge was allowed to range from 0.5 to 1.5 inches per year (in/yr), which is within plus/minus 50 percent of the value applied to most of the JGWFS model.
- Dominguez Channel conductance was allowed to range within 10 percent (log-transformed) of the values used in for the JGWFS model.
- GHB conductance was calculated for each GHB using the following equation: GHB conductance = K*L*W/M, where K is the horizontal hydraulic conductivity at a particular cell, L is the thickness of the cell, W is the size of the cell in the direction perpendicular to flow, and M is half the size of the cell in the direction parallel to flow.
- GHB heads were assigned using six to nine pilot points per aquifer layer including one point at each model corner (i.e., four points), one point in the middle of the northeast and southwest boundaries of the model (i.e., two points), and additional points at the locations where water-level data were available near the boundary. Aquitard GHB heads were assigned as averages of heads in the underlying and overlaying aquifers. For example, the GHB head in Layer 11 of the Gage-Lynwood Aquitard (GLA) was calculated as the average of head in the Gage Aquifer (Layer 9) and in the Lynwood Aquifer (Layer 13). GHB head in GLA Layer 10 was then calculated as the average of heads in Layers 9 and 11.

Parameter Limits for Transport Parameters

- Porosity was allowed to range from 10 to 50 percent (JGWFS value plus/minus 66 percent).
- Longitudinal dispersivity (D_L) was allowed to range from 0.1 to 1,000 feet, based on the scale of the plume. The JGWFS model used a small value, due to its use of a stochastic hydraulic conductivity field, which introduces dispersion via variation at the model cell scale.
- Transverse (D_T) to longitudinal dispersivity ratio was allowed to range from 0.001 to 1, based on literature estimates.
- Vertical (D_Z) to longitudinal dispersivity ratio was allowed to range from 1x10-6 to 1, based on literature estimates.
- Kd was allowed to range from 0.001 to 6 milliliters per gallon (mL/g), based on field and literature data.
- R was calculated by MT3DMS from field-measured bulk density values and Kd as estimated by PEST. The allowable range was 1 to 90.

- Source concentrations for both p-CBSA and chlorobenzene were allowed to range from 0.0001 to 100,000 micrograms per liter (µg/L). The high end of this range was lower than appropriate, due to a data entry error. The low end was designed to allow the flexibility to account for the coarseness with which the source geometry had to be defined, given the cell size in the source area (i.e., the source had to be no smaller than a single grid cell). The concentration ranges for the chlorobenzene and p-CBSA sources were revised to 0.001 to 500,000 µg/L and 0.001 to 1,000,000 µg/L, respectively, to account for the solubility limits of these constituents.
- Source start time was assumed to be January 1945.

Preferred Values for Model Parameters

As discussed in Section 2.1, a "preferred condition" for each calibration parameter was defined by the regularization equations as a set of preferred parameter values, from which deviations are tolerable only to the extent that they are supported by the data. The preferred values for model calibration parameters and the rationale for these values are presented in Table 3-3.

TABLE 3-3
Preferred Values for Model Parameters

Parameter	Location	Preferred Value	Units	Rationale
Flow				
River Conductance	Dominguez Channel	11,000	ft^2/d	Same as JGWFS model
Recharge		1	in/yr	Same as majority of JGWFS model
Horizontal Hydraulic Conductivity	UBF	1.9	ft/d	Geometric mean of aquifer test data in this HSU
	MBFB	20	ft/d	Geometric mean of aquifer test data in this HSU
	MBFM	0.015	ft/d	Same as JGWFS model
	MBFC	145	ft/d	Geometric mean of aquifer test data in this HSU
	LBF	0.005	ft/d	Professional judgment based on previous calibration results
	Gage	36	ft/d	Geometric mean of aquifer test data in this HSU
	GLA	0.0064	ft/d	Same as JGWFS model
	Lynwood	113	ft/d	Geometric mean of aquifer test data in this HSU
Horizontal/Vertical Hydraulic Conductivity Ratio	UBF	1,000		Professional judgment based on previous calibration results
	MBFB	100		Professional judgment based on previous calibration results
	All others (fixed)	10		Generally the same as JGWFS model
Transport				
Porosity	All layers	0.3		Same as JGWFS model
Longitudinal dispersivity	All layers	100	ft	Based on scale of chlorobenzene plume
Transverse to longitudinal dispersivity ratio	All layers	0.5		Same as JGWFS model
Vertical to longitudinal dispersivity ratio	All layers	0.001		Same as JGWFS model
Chlorobenzene source concentration	MBFB	350,000	µg/L	Same as JGWFS model
	MBFM	350,000	µg/L	Same as JGWFS model
	MBFC	15,000	µg/L	Same as JGWFS model
	LBF	10,000	µg/L	Same as JGWFS model
	Gage	7,000	µg/L	Same as JGWFS model
Chlorobenzene Kd	UBF	0.0053	mL/g	Same as JGWFS model
	MBFB	0.039	mL/g	Same as JGWFS model
	MBFM	0.018	mL/g	Same as JGWFS model
	MBFC	0.13	mL/g	Same as JGWFS model
	LBF	0.43	mL/g	Same as JGWFS model
	Gage	0.27	mL/g	Same as JGWFS model
	GLA	0.73	mL/g	Same as JGWFS model
	Lynwood	0.53	mL/g	Same as JGWFS model
p-CBSA source concentration	MBFB	500,000	µg/L	Same as JGWFS model

3.2 Calibration Targets

Five main groups of calibration targets were selected for initial calibration. These groups include (1) the Fall 1995 water level data (heads), (2) head differences based on the Fall 1995 water level data, (3) all available chlorobenzene concentration data (i.e., from 1983 through 2002), (4) all available p-CBSA concentration data, and (5) chlorobenzene and p-CBSA mass targets estimated based on kriging of concentration data. The calibration targets were subgrouped by model layers (1, 2, 4, 5, 9, and 13).

The weight of each group was initially set to result in an approximately equal contribution from each subgroup to the objective function. These weights were further adjusted during the calibration process to improve the match between the observed and simulated results and focus calibration on the most pertinent aspects with regard to the modeling objectives, such as the above-maximum-contaminant-level (MCL) chlorobenzene distribution in the MBFC and the Gage Aquifer. The groups of calibration targets and weighting of these targets in the calibration process are discussed below.

3.2.1 Heads

All available water level data for Fall 1995 were used as calibration targets. If several head measurements were available for the same well, an average of these measurements was used as a calibration target for this location. Two head calibration targets located in an inactive portion of Layer 1 (wells MW2AG and MW07) were assigned to Layer 2 in order to include these measurements in the objective function. It is important to note that there is a set of wells (mostly water table wells) that was used for the JGWFS model calibration; however, the data source for these wells was not available. While these data were used in the initial calibration, the usability of these data for RD model development will need to be assessed. All heads were assigned equal weights in the calibration process. Note that the 2004 baseline water level measurements were not available at the time of the initial calibration and assessment of model uncertainty. Because the 2004 water levels are generally consistent with the Fall 1995 data, it is not likely that using the 2004 water levels for the steady state flow calibration would change the outcome of the model uncertainty assessment. However, the 2004 data will be used for the final RD model.

3.2.2 Vertical Head Differences

Vertical head difference calibration targets were calculated for 38 locations where a well was located near another well screened in a different hydrostratigraphic unit. The distance between wells selected for calculating vertical head differences was limited to 20 feet. Given an average lateral gradient of 0.001 foot per foot (ft/ft) in the model domain, the distance of 20 feet would result in a maximum error of only 0.02 foot in estimates of vertical head differences. This amount of error is negligible given the average head difference of 2.62 feet in the 38 well pairs. Head differences were weighted by the inverse of the absolute value of the head difference, to equalize the importance of small and large head differences.

3.2.3 Chlorobenzene Concentrations

All available chlorobenzene concentrations (i.e., for years 1983 through 2002) were used as calibration targets. Note that the 2004 baseline sampling results were not available at the

time of the initial baseline calibration and assessment of model uncertainty for predictions of mass and volume reduction. These results were added to calibration targets during the revisions to the baseline calibration (see Section 4.2.2.1) and were used for simulating containment failure scenarios. Because the 2004 results are generally consistent with the historical data, it is not likely that adding these results to calibration targets would change the outcome of the model uncertainty assessment with regard to mass and volume reduction. Therefore, this assessment was not repeated after the 2004 data became available.

The weights of chlorobenzene concentration calibration targets have been adjusted several times in the process of calibration to achieve a better match between the observed and simulated distributions. Initially, concentrations were weighted by the inverse of an assumed log-transformed standard deviation of 0.05 (based on a visual examination of relatively stable portions of chemographs at the site). The standard deviation for a small concentration was assumed to be an order of magnitude less than that for a concentration 10 times higher, which resulted in a weighting scheme that values nondetects and low concentrations approximately as much as high concentrations. This distribution of weights resulted in a good match of nondetects, but marginal match of the high-concentration plume. A linear weighting scheme was then assumed, which resulted in an improved match in wells with high chlorobenzene concentrations. The weights were further increased in several wells located at and within the 1,000-µg/L concentration contour to better reproduce the distribution of chlorobenzene above 1,000 µg/L.

3.2.4 p-CBSA Concentrations

All available p-CBSA concentrations (i.e., for years 1983 through 2002) were used as calibration targets. Concentrations of p-CBSA were weighted similarly to those of chlorobenzene.

3.2.5 Mass Targets

Mass targets for chlorobenzene were developed based on kriging of available concentration data for a given year (i.e., for years 1983 through 2002). The kriged concentration distribution at a particular time was summed to produce a particular mass target. All mass targets for a particular chemical had equal weights.

3.3 Calibration Results

This section describes the results of the baseline calibration including flow and transport calibration, contribution to objective function, and calibrated distributions of model parameters.

3.3.1 Flow Calibration

The results of the flow calibration are presented in Figures 3-1 through 3-3. Figure 3-1 presents a scatter diagram of simulated versus measured water levels (heads) and calibration statistics.

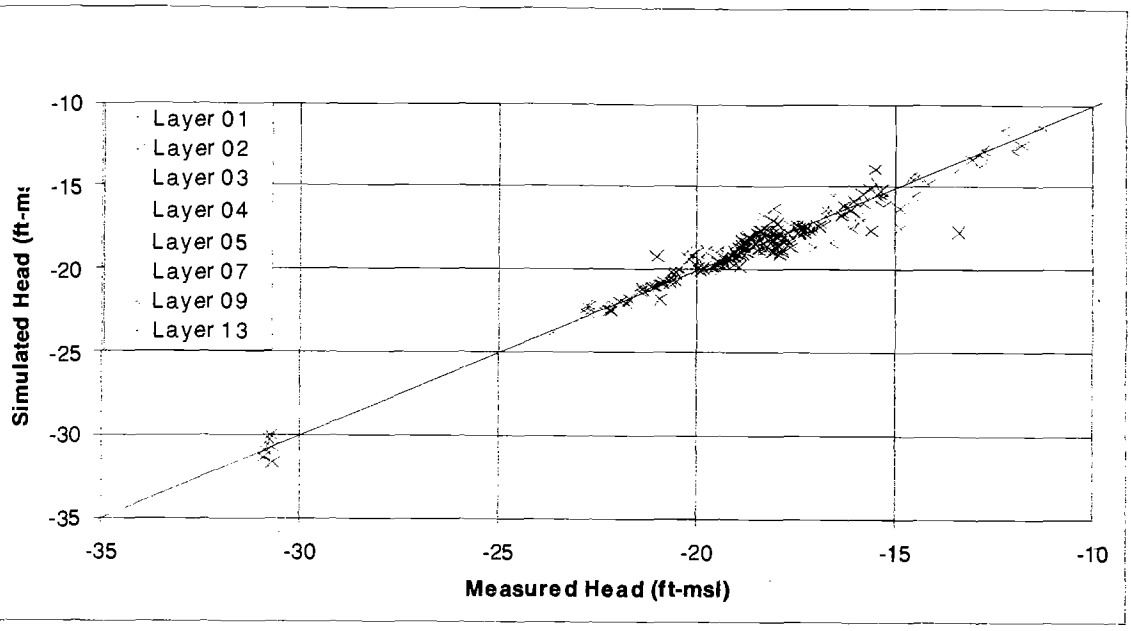
Figure 3-2 presents simulated water level contours and measured water levels for model Layers 2, 5, and 9 representing the water table, MBFC, and Gage Aquifer. Figure 3-3 presents simulated versus observed vertical head differences.

These figures demonstrate agreement between the observed and simulated water levels and vertical head differences. The scatter plot of measured and simulated water levels shown in Figure 3-1 has a slope that is similar to the line of perfect agreement (i.e., 1:1 slope), and is located relatively close to this line, indicating a good agreement between the simulated and measured heads in all units.

The calibration error, as measured by the root mean squared (RMS) of simulated heads versus measured water level elevations, is 0.68 foot when data for all 215 monitoring wells are considered; this also indicates a good match between observed and simulated water levels.

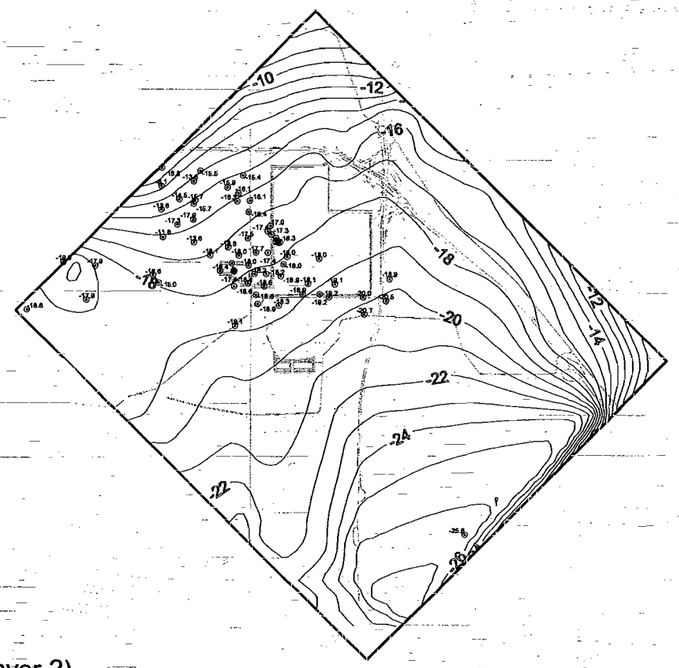
The scatter plot of measured and simulated vertical head differences shown in Figure 3-3 is also located close to the line of perfect agreement, indicating good agreement between the simulated and measured vertical head differences. The use of vertical head differences between the model layers as calibration targets allowed better estimates of vertical hydraulic conductivities than using water levels alone.

FIGURE 3-1
Simulated vs. Measured Water Levels – Baseline Calibration

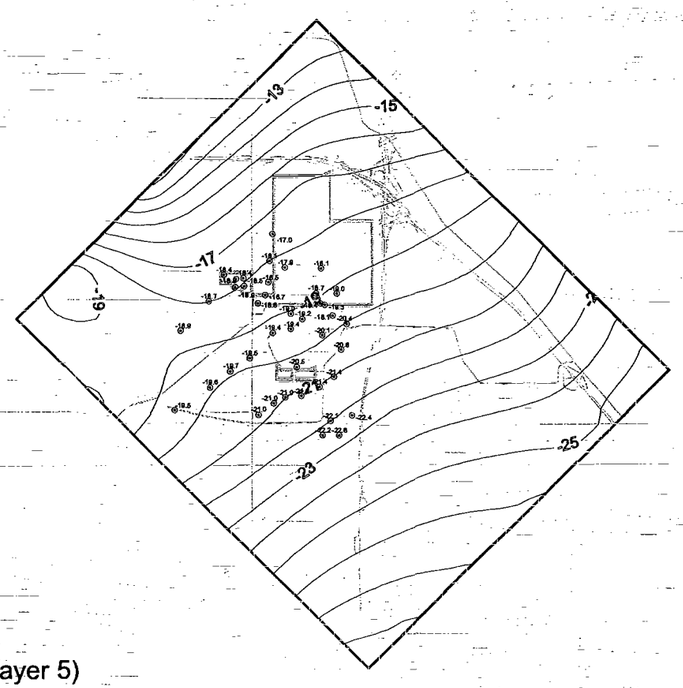


Head Calibration Statistics

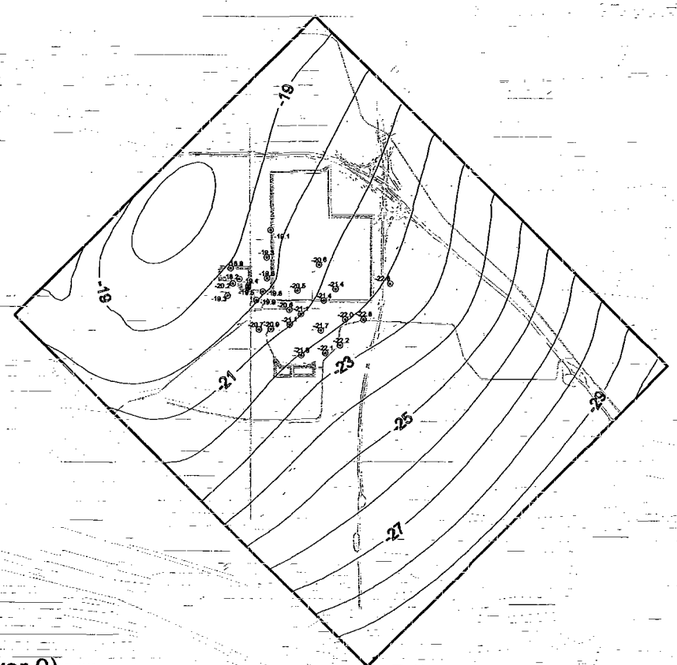
Count	215
Min	-1.84
Max	4.32
Mean	0.00
Absolute Mean	0.46
RMS	0.68
RMS/Range	3.5%



MBFB (Model Layer 2)



MBFC (Model Layer 5)



Gage (Model Layer 9)

- Calibration Well Locations with Measured October 1995 Water Levels (ft msl)
- Simulated October 1995 Water Level Contours (ft msl)
- Streets
- Site Location
- Model Domain
- Base Map

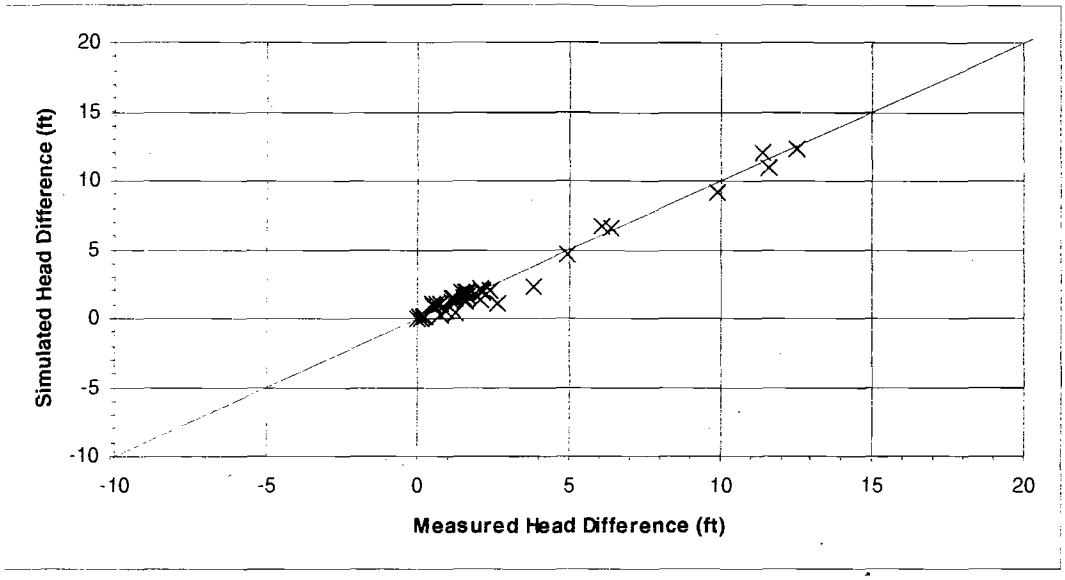
N

5000 Feet 5000 Feet

Figure 3-2
Simulated 1995 Water Level Contours
Baseline Calibration

J:\Montrose_GW_Remedy\plots\092904_report_plot1

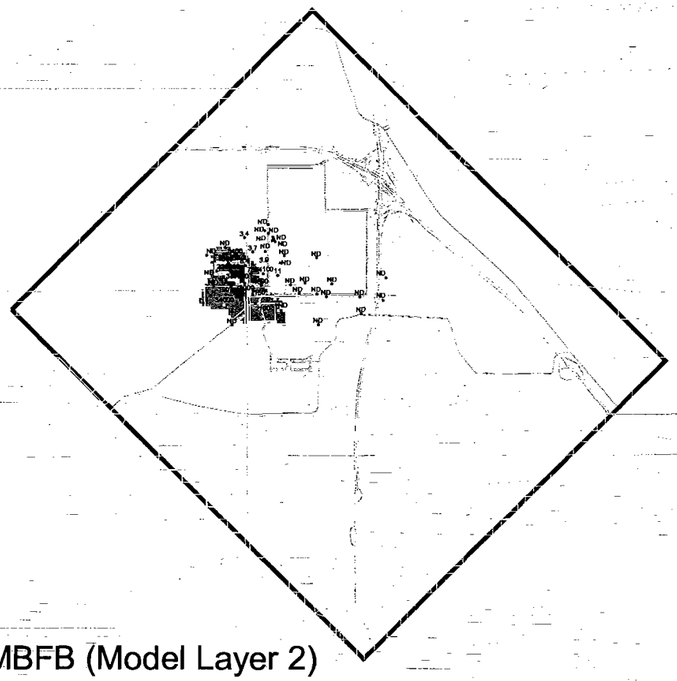
FIGURE 3-3
Simulated vs. Measured Vertical Head Differences



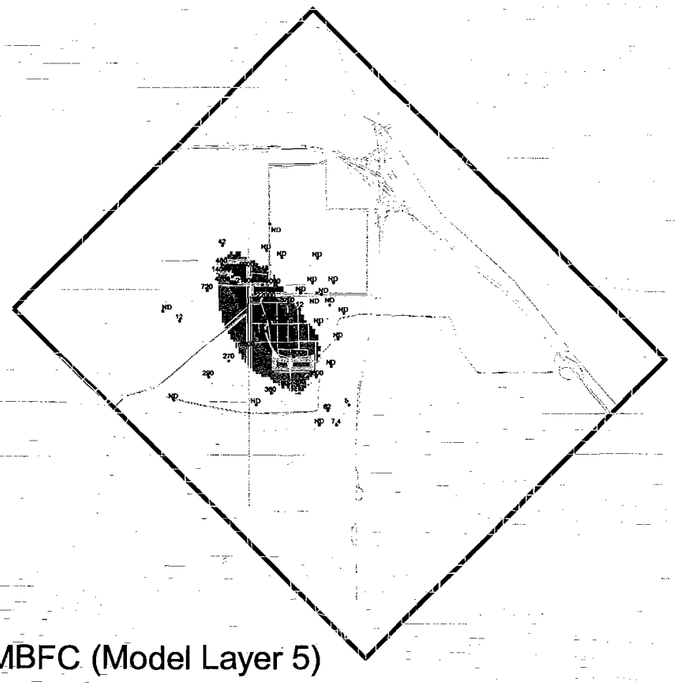
3.3.2 Transport Calibration

The results of transport calibration are presented in Figures 3-4 and 3-5 for chlorobenzene and p-CBSA. These results indicate a reasonably good match between measured and simulated concentrations of both constituents. The model reproduces the observed difference in the orientation of the plumes in the MBFC and the Gage Aquifer. The simulated chlorobenzene and p-CBSA plumes in the Gage Aquifer are oriented slightly more to the east compared to the plumes of these constituents in the MBFC, which is consistent with field data. Low to nondetectable concentrations of chlorobenzene observed in Gage monitoring wells G-15 and G-16 underneath the high concentration plume in the MBFC were also reproduced by the model.

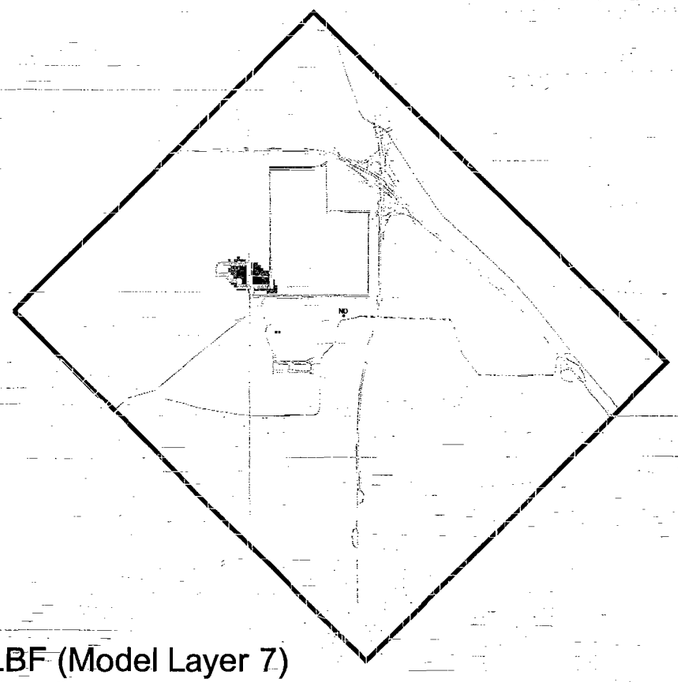
As discussed in the calibration targets section (Section 3.2) of this report, the baseline transport calibration was focused on reproducing the high (i.e., above 1,000 µg/L) concentrations of chlorobenzene, because the distribution of the high-concentration plume has the greatest effect on the design of the remedial wellfield. A linear weighting scheme was generally assumed for most concentration targets (i.e., equal weights for low and high concentrations) with the exception of several wells at the toe of the 1,000-µg/L plume (BF-16, BF-11, and BF-17), for which the weights were increased to better reproduce the downgradient extent of the plume.



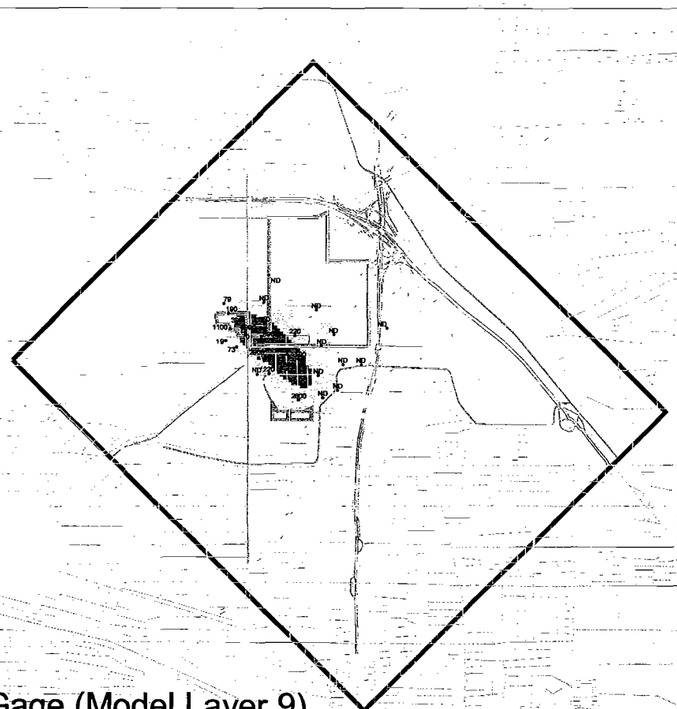
MBFB (Model Layer 2)



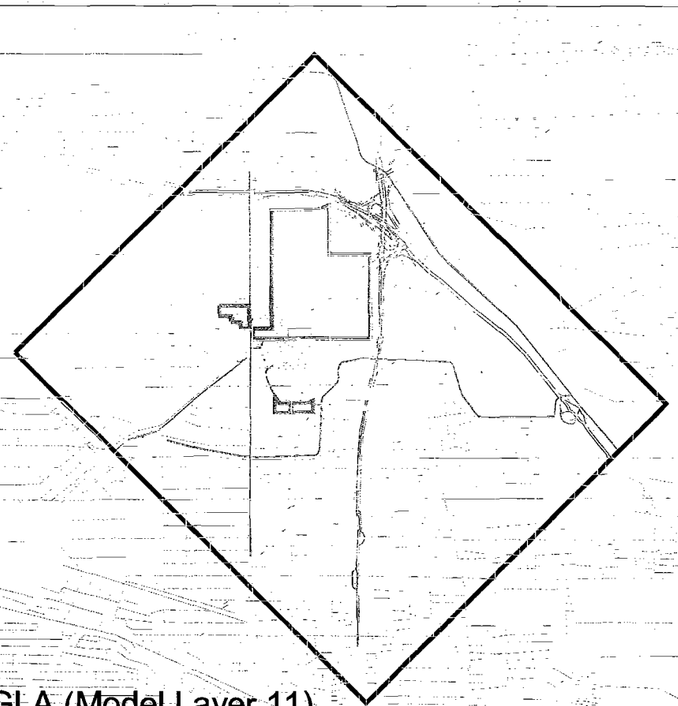
MBFC (Model Layer 5)



LBF (Model Layer 7)



Gage (Model Layer 9)



GLA (Model Layer 11)

· Observed Chlorobenzene Concentrations (µg/L)

Streets
Site Location
Model Domain
Base Map

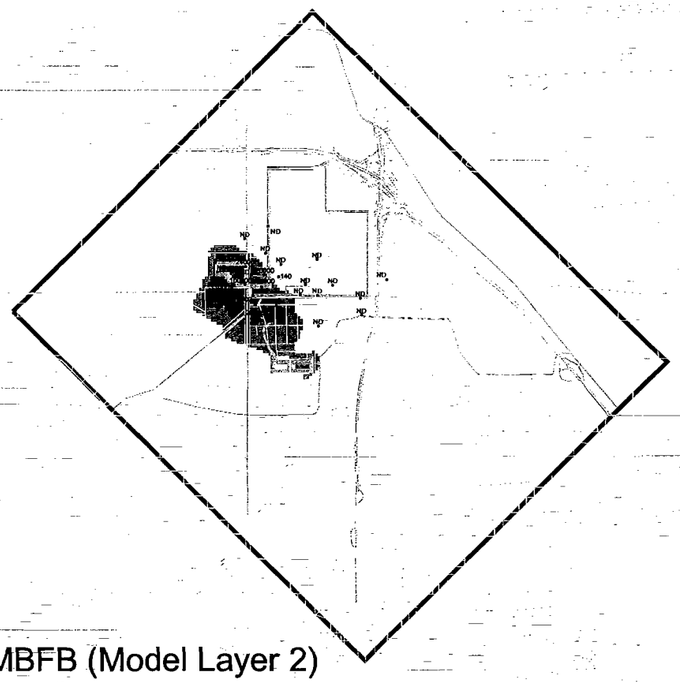
Color Codes for Simulated Chlorobenzene Concentrations (µg/L)
70 - 1000
1000 - 5000
5000 - 10000
10000 - 100000
100000 - 1000000

N

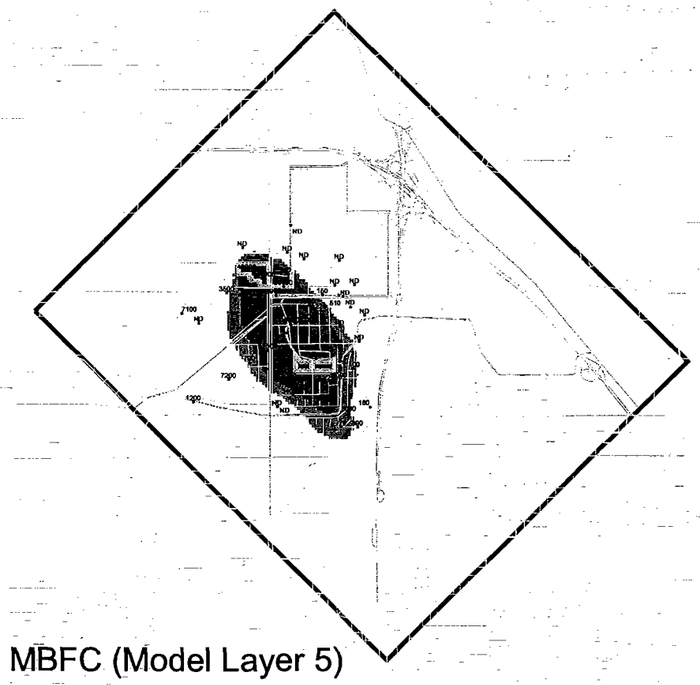
5000 Feet 5000 Feet

Figure 3-4
Simulated Historical
Chlorobenzene Distributions (1945-1995)
Baseline Calibration

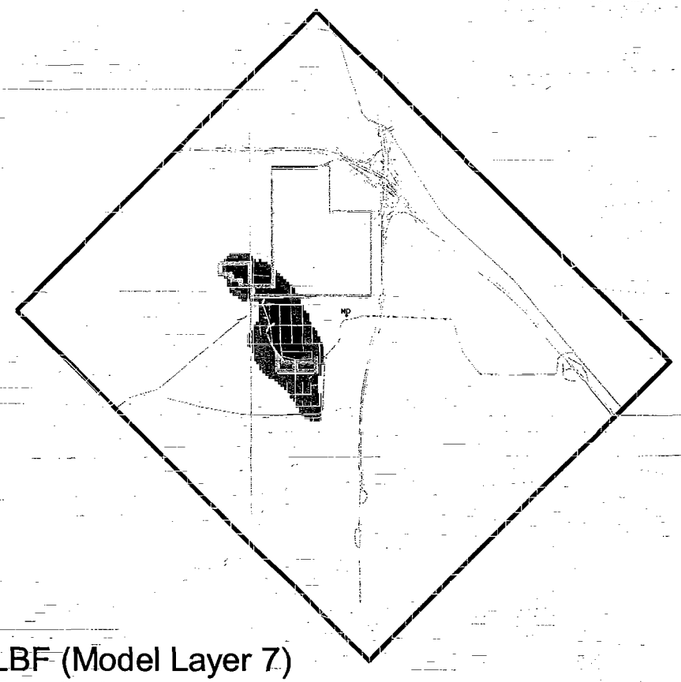
J:\Montrose_GW_Remedy\Plots\092904_report_plot2



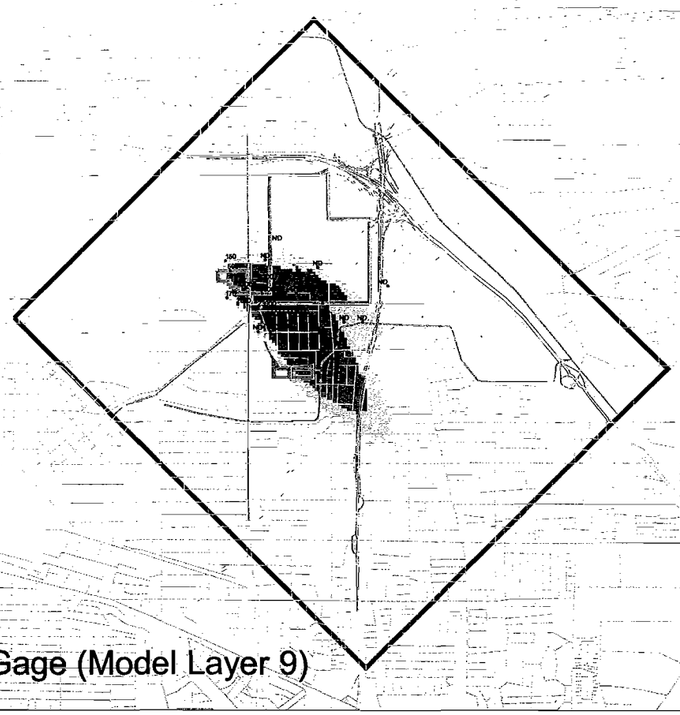
MBFB (Model Layer 2)



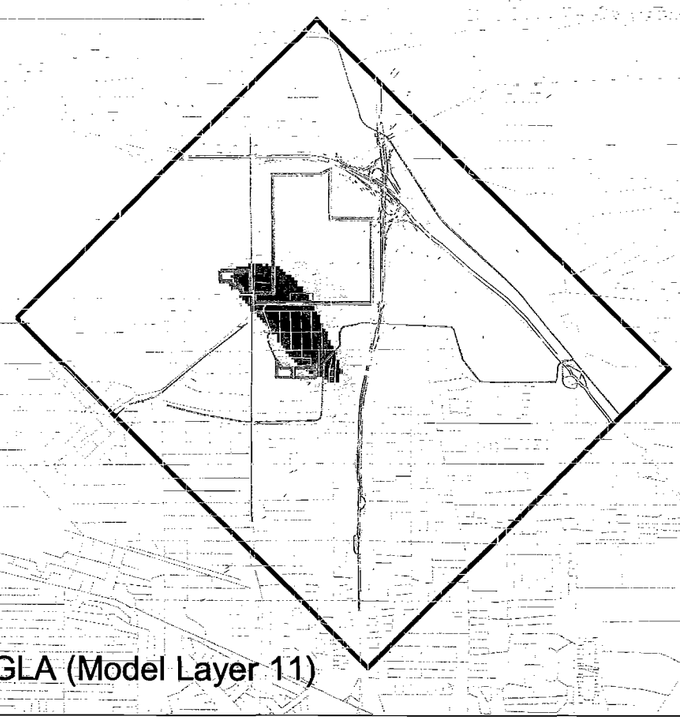
MBFC (Model Layer 5)



LBF (Model Layer 7)



Gage (Model Layer 9)



GLA (Model Layer 11)

- Observed p-CBSA Concentrations (µg/L)

Streets
Site Location
Model Domain
Base Map

Simulated p-CBSA Concentrations (µg/L)
200 - 1000
1000 - 5000
5000 - 10000
10000 - 100000
100000 - 1000000

N

5000 Feet 5000 Feet

Figure 3-5
Simulated Historical
p-CBSA Distributions (1945-1995)
Baseline Calibration

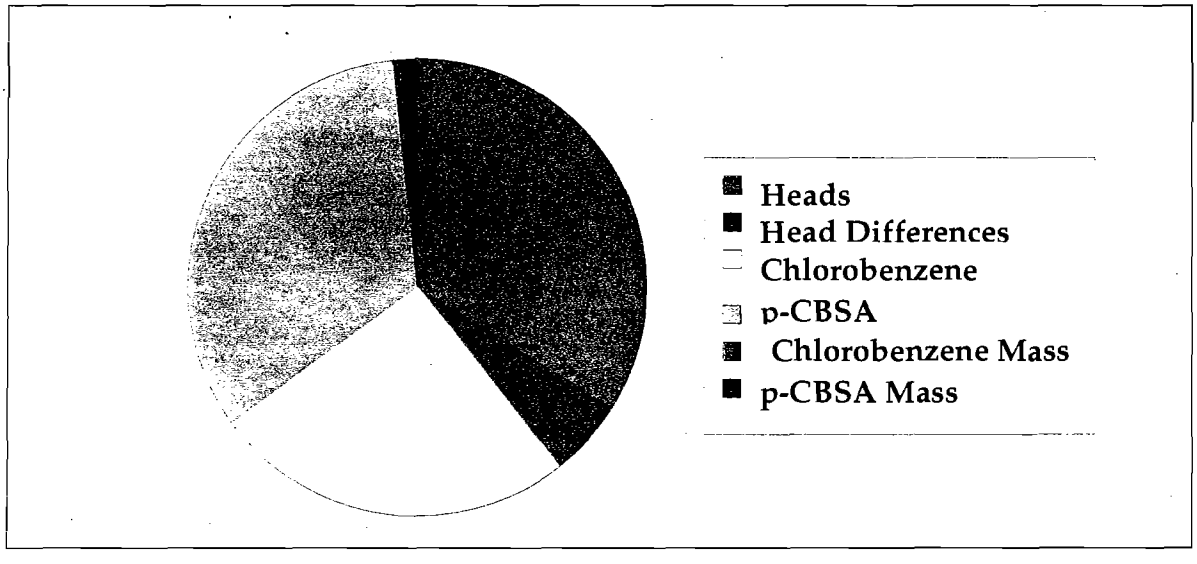
J:\Montrose_GW_Remedy\Plots\092804_report_plot3

The primary difficulty of transport calibration was to reproduce the steep concentration gradient at the toe of the chlorobenzene plume, such as, changes in chlorobenzene concentrations from 8,800 µg/L in well BF-11 to 21 µg/L in well BF-25 (located 1,250 feet downgradient of BF-11). These difficulties may be attributed to uncertainties associated with the historical flowfield, timing and strength of contaminant sources, and the simplification of these parameters in the calibration process. As discussed above, the baseline calibration assumed steady-state flow based on 1995 water levels, and constant-strength sources. The actual historical flowfield likely varied with time. The source strength also likely varied with time in response to changes in recharge, subsurface flow through the source area, and other unknown conditions. As a result, simulated concentration gradients at the toe of the chlorobenzene plume are less steep than observed. Simulated concentrations in high-concentration wells (BF-16, BF-11, and BF-17) are slightly underestimated, while concentrations in the downgradient low-concentration wells are overestimated. However, the overall simulated distributions of both p-CBSA and chlorobenzene match the observed conditions reasonably well.

3.3.3 Contribution to Objective Function

The contribution to the objective function from different calibration targets is presented in Figure 3-6. Calibration weights were designed such that heads, chlorobenzene concentrations, and p-CBSA concentrations would contribute in fairly equal proportion to the objective function. Vertical head differences and mass targets were considered secondary targets. As a result, the baseline calibration accounts for a number of measured and estimated parameters and is considered to be a reasonable representation of flow and contaminant transport conditions beneath the Site.

FIGURE 3-6
Contribution to Objective Function from Calibration Targets



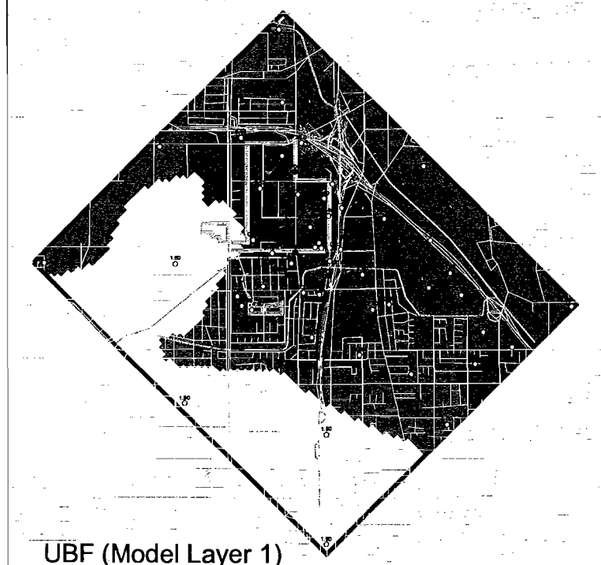
3.3.4 Calibrated Distributions of Model Parameters

The calibrated distribution of horizontal and vertical hydraulic conductivity is presented in Figures 3-7 and 3-8. Note that for the HSUs represented by multiple model layers such as MBFC, LBF, and GLA, the distribution of these properties is the same in all layers representing a given unit, and only the top layer is shown on the figures. The calibrated recharge distribution is shown in Figure 3-9.

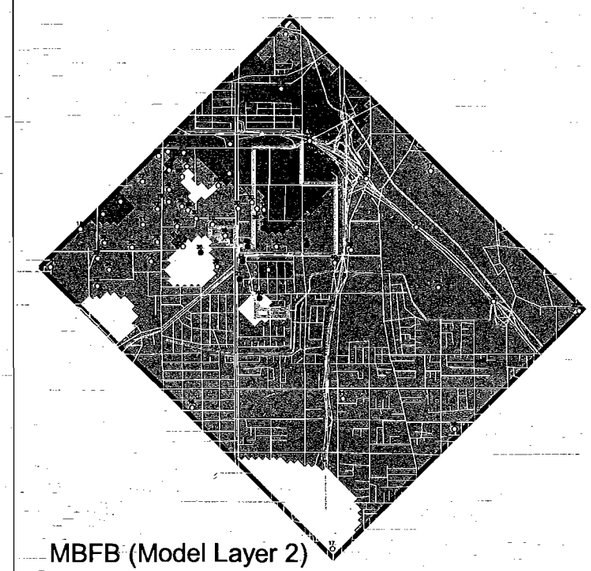
Calibrated transport parameters including porosity; longitudinal, transverse, and vertical dispersivity (D_L, D_T, and D_Z); Kd; R; bulk density; and chlorobenzene and p-CBSA source concentrations are presented in Table 3-3.

TABLE 3-4
Calibrated Transport Parameters

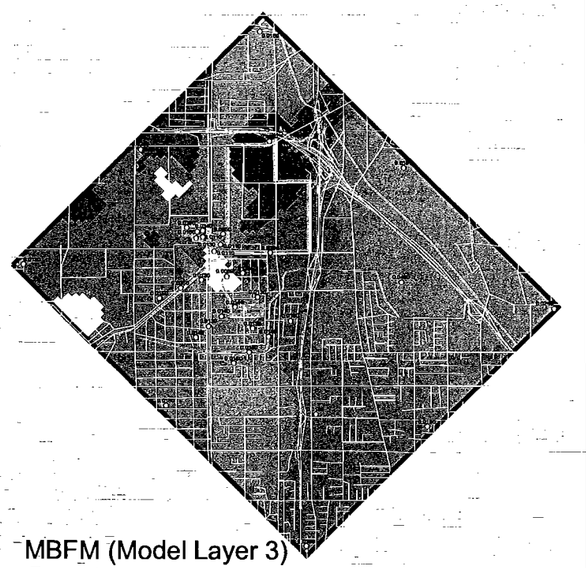
HSU	Porosity	D_L (ft)	D_T (ft)	D_z (ft)	Chlorobenzene Source (mg/L)	Chlorobenzene Kd (mL/g)	R (n_{eff} = n)	p-CBSA Source (mg/L)	Bulk Density (g/cm^3)
UBF	0.33	110	0.5	0.001		0.0053	1.02		1.49
MBFB	0.30	330	0.29	0.0056	1,000,000	0.069	1.34	1,000,000	1.49
MBFM	0.30	24			380,000	0.019	1.08	1,000,000	1.25
MBFC	0.27	25	1	0.0005	240,000	0.036	1.21	1,000,000	1.59
LBF	0.30	7.5	0.5	0.0007	11,000	0.27	2.37	250,000	1.52
Gage	0.10	14	0.34	0.001	7,100	0.067	2.03	45,000	1.53
GLA	0.30	59	0.5	0.001		0.73	4.6		1.51
Lynwood	0.31	100	0.49	0.001		0.052	1.30		1.76



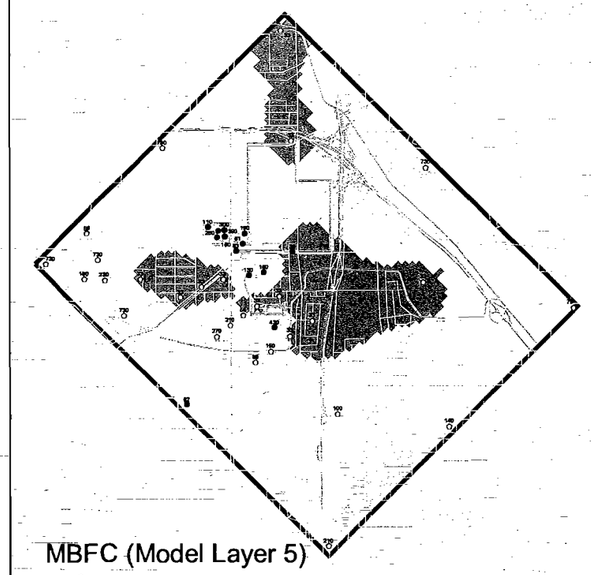
UBF (Model Layer 1)



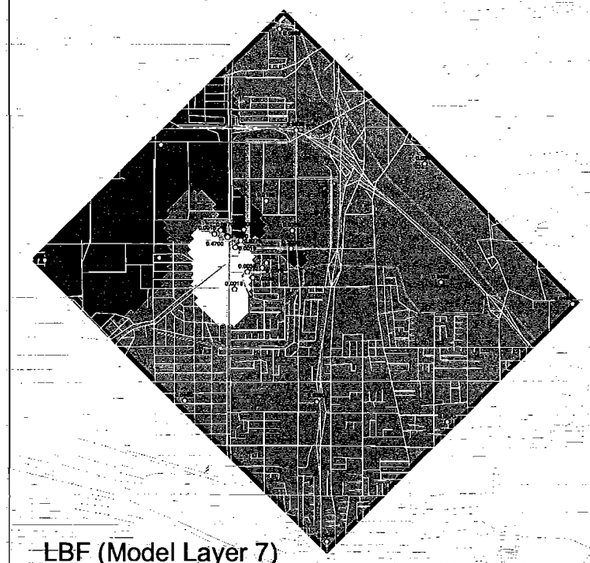
MBFB (Model Layer 2)



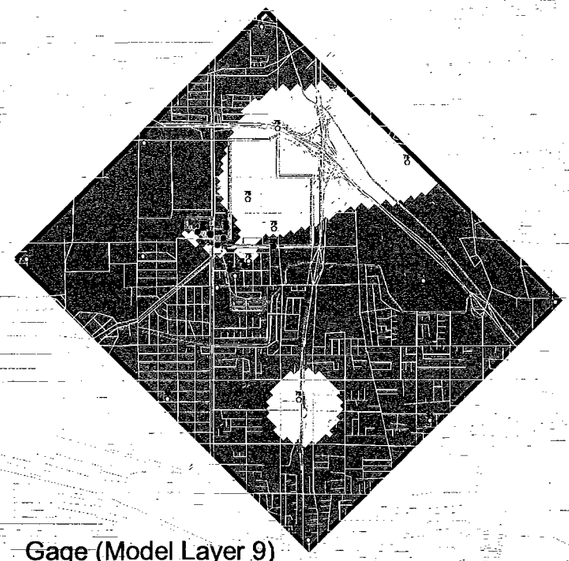
MBFM (Model Layer 3)



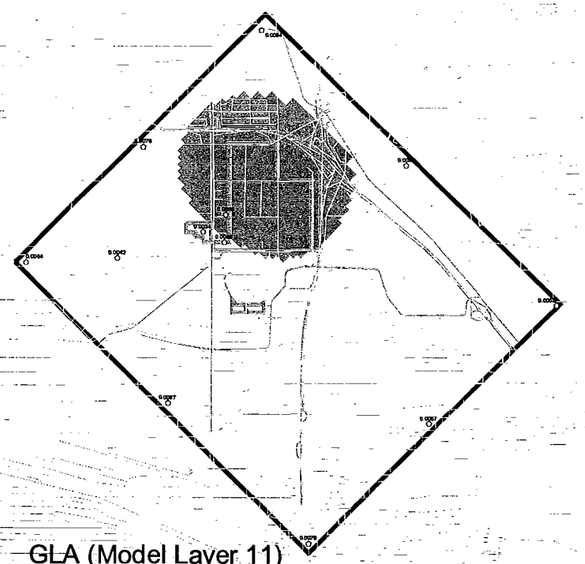
MBFC (Model Layer 5)



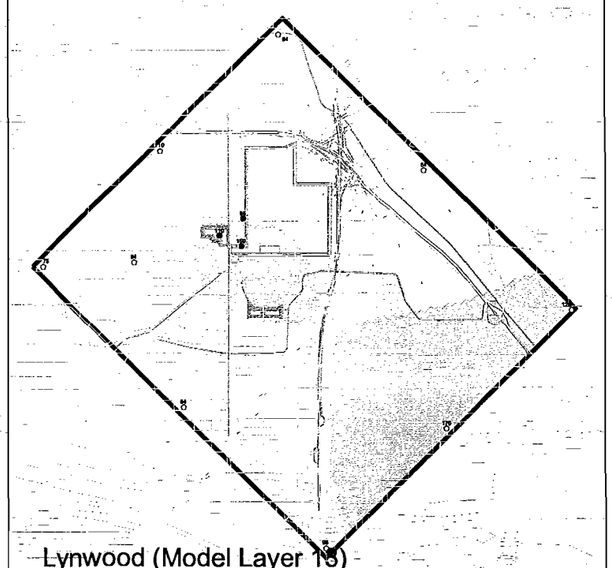
LBF (Model Layer 7)



Gage (Model Layer 9)



GLA (Model Layer 11)



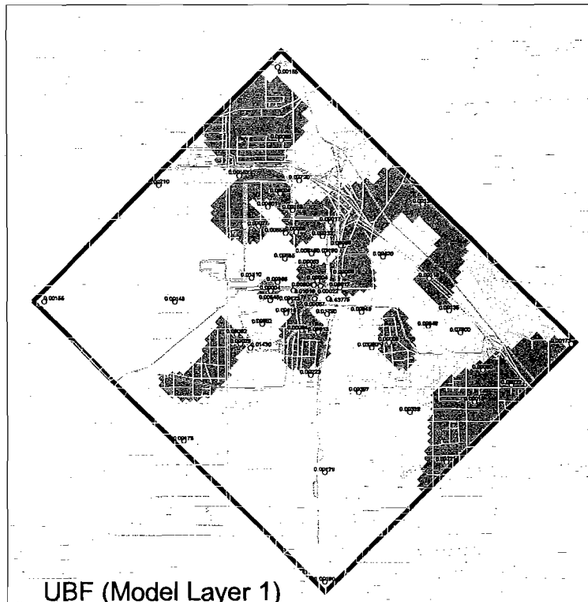
Lynwood (Model Layer 13)

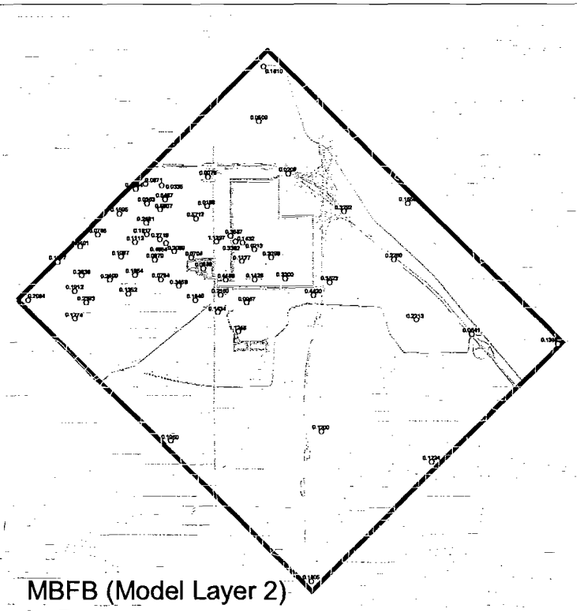
- ◦ Pilot Points, Values from Model Calibration
- • Pilot Points, Values from Aquifer Tests
- Streets
- Site Location
- Model Domain
- Base Map

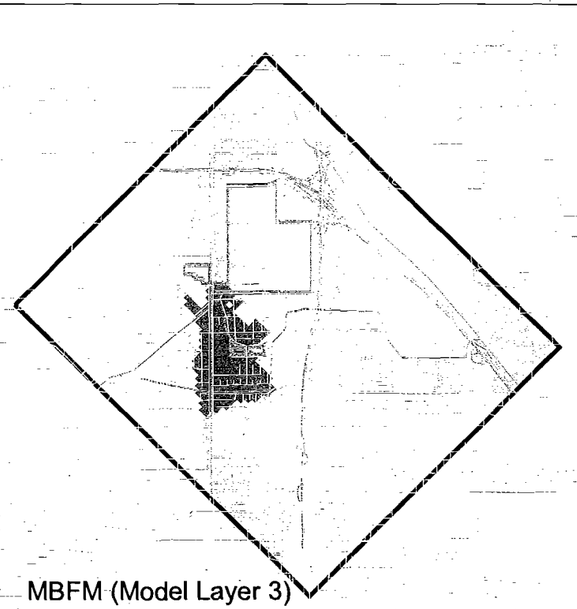
Color Codes for Horizontal Hydraulic Conductivity (ft/day)

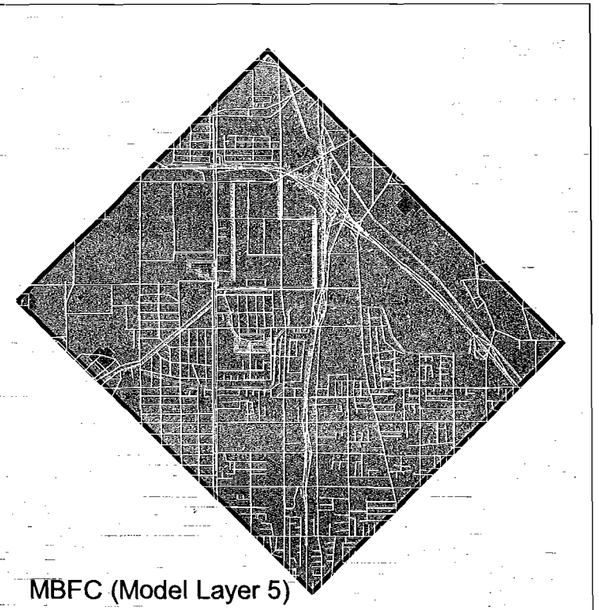
0.001 - 0.01	50 - 100
0.01 - 0.1	100 - 300
0.1 - 1	300 - 500
1 - 10	500 - 1000
10 - 50	

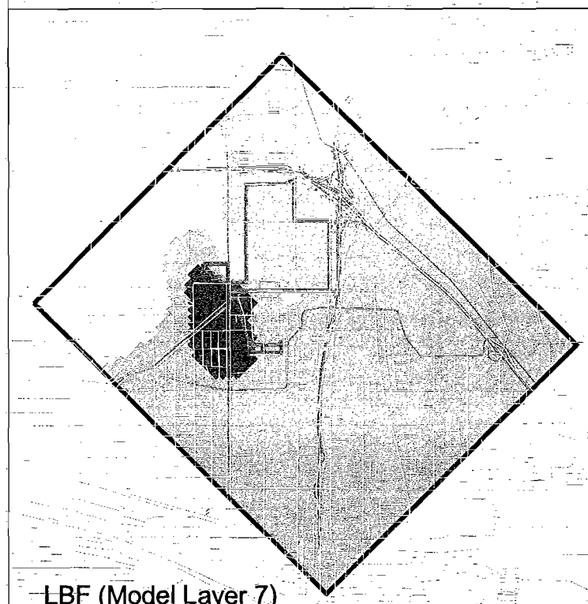
Figure 3-7
Calibrated Horizontal Hydraulic Conductivity
Baseline Calibration


UBF (Model Layer 1)

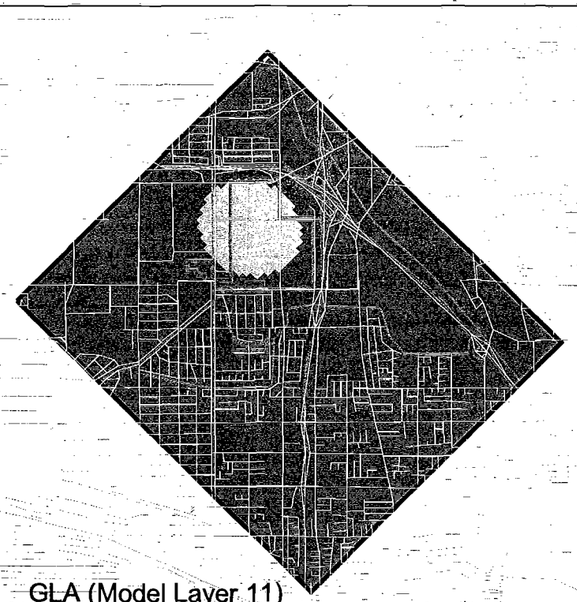

MBFB (Model Layer 2)

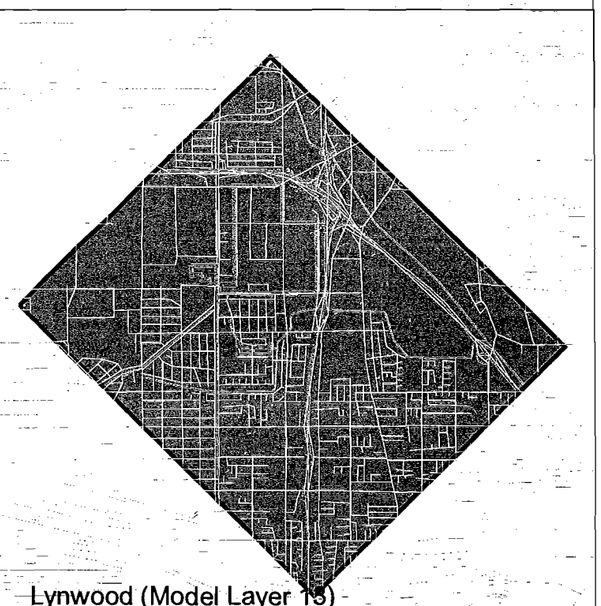

MBFM (Model Layer 3)


MBFC (Model Layer 5)


LBF (Model Layer 7)


Gage (Model Layer 9)


GLA (Model Layer 11)


Lynwood (Model Layer 13)

- Pilot Points, Values from Model Calibration
- Pilot Points, Values from Aquifer Tests
- Streets
- Site Location
- Model Domain
- Base Map

Color Codes for Vertical Hydraulic Conductivity (ft/day)

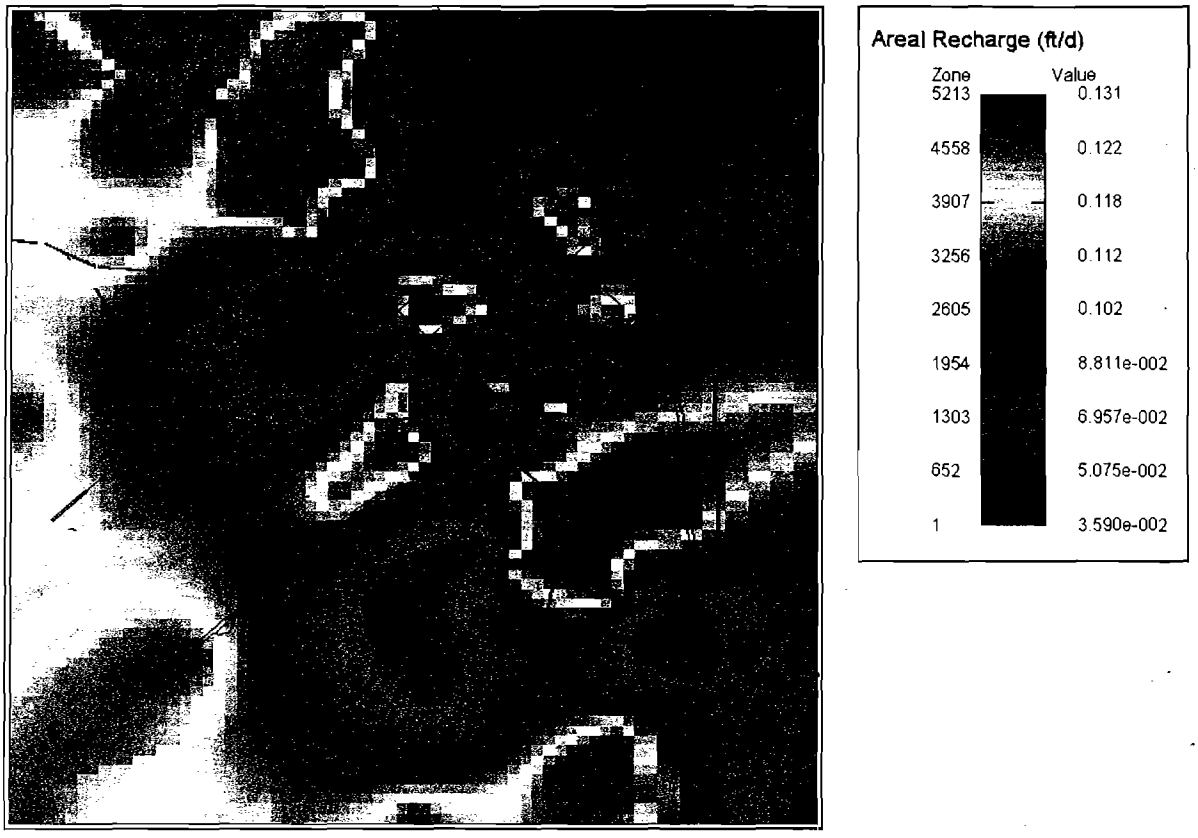
- <0.001
- 0.001 - 0.01
- 0.01 - 0.1
- 0.1 - 1
- 1 - 10
- 10 - 50
- 50 - 100
- 100 - 300
- 300 - 500

N

6000 Ft 6000 Ft

Figure 3-8
Calibrated Vertical Hydraulic Conductivity
Baseline Calibration

FIGURE 3-9
Calibrated Recharge



The values and distribution of model parameters obtained as the result of the baseline calibration are reasonable for the hydrogeologic system beneath the Site. However, as discussed above, this combination of model parameters resulting in a good match between the observed and simulated conditions is one of many possible and equally reasonable combinations, which could be obtained with PEST calibration. Consequently, other combinations of model parameters should be considered to assess the range of possible calibration solutions and the impact of these solutions on model predictions. Alternative calibration solutions and their effect on model predictions are discussed in Section 4.

4. Data Gap Analysis

As discussed above, predictive calibration was used to assess the predictive uncertainty of the model and identify the types of data that have the most effect on reducing this uncertainty. Predictive calibration involves the use of "predictive targets" in the calibration process in addition to the calibration targets. The model is run by PEST in both calibration and predictive modes, and the objective function is minimized by reducing residuals associated with both calibration and predictive targets. The predictive targets used for this data gap analysis were based on the standards and requirements of the ROD (EPA, 1999) and included (1) rates of contaminant volume and mass reduction in the aquifer targets, and (2) plume containment targets (see Sections 4.1 and 4.2, respectively).

The calibration portion of the model run during predictive calibration was similar to the baseline calibration (see Section 3). The predictive portion of the run was performed using steady-state flow and transient transport for a 50-year period. For the purposes of this analysis, the initial conditions for the chlorobenzene plume were based on the Fall 1995 data (same data set used for the JGWFS). The more recent data set (i.e., 2004 baseline data) will be used for the final design. For the predictive portion of the run, it was assumed that the sources of chlorobenzene were contained in all units and did not contribute to the chlorobenzene concentrations. Consequently, the source terms were not simulated in the predictive portion of the model run.

The predictive run simulated the remedial wellfield that was developed for Scenario 5 of the JGWFS (CH2M HILL, 1998) and revised as part of the preliminary groundwater modeling (Hargis, 2003). The use of this remedial wellfield was intended only for the purposes of the data gap analysis. The final remedial wellfield will be developed as part of the remedial design and optimization task (RD) using the revised RD model.

As discussed in Section 2.2, the identification of the most critical parameters affecting the performance of the remedial wellfield was performed based on several qualitative and quantitative analyses. These analyses included the qualitative comparison of calibrated model parameter distributions resulting in different predictive outcomes, evaluation of the sensitivity of the predictive targets to each model parameter, and evaluation of the contribution of each parameter to the predictive uncertainty of the model.

For this data gap analysis, predictive uncertainty of the model was evaluated only for the selected ROD standards and performance criteria mentioned above, such as contaminant mass and volume reduction and containment of the plume. However, all standards and requirements of the ROD will be considered for the design of the remedial wellfield. The discussion of the predictive uncertainty analysis pertaining to (1) the contaminant mass and volume reduction rates, and (2) hydraulic containment of the contaminant plume is presented in Sections 4.1 and 4.2, respectively.

4.1 Assessment of Model Uncertainty for Predictions of Mass and Volume Reduction

The ability of the remedial wellfield to clean up the chlorobenzene plume in a reasonable timeframe is an important requirement of the ROD. The cleanup rates could be expressed as the contaminant mass and volume reduction numerical targets. The predictive uncertainty of the model with regard to the mass and volume reduction was assessed to identify model parameters that are the most critical for reducing this uncertainty. The mass and volume reduction targets, calibration results and predictions of mass and volume reduction for the best and worst case, and the model uncertainty for predictions of mass and volume reduction are discussed below.

4.1.1 Mass and Volume Reduction Targets

Predictive targets for mass and volume reductions were developed based on the requirements of the ROD and the results of model simulations performed for the JGWFS. The volume and mass reduction rates were estimated as percent volume and mass reduction of the plume within the 70-µg/L contour (i.e., above-MCL distribution), based on the 1990 through 1995 chlorobenzene concentration data (the same data set as was used for the JGWFS). This approach is similar to that of the JGWFS with the exception that, for the purposes of this data gap analysis, all of the above-MCL chlorobenzene plume including the dense nonaqueous-phase liquid (DNAPL)-impacted area was considered for the estimates of the plume volume and mass. For the JGWFS, the DNAPL-impacted area was excluded from the calculations of the mass and volume of the plume, because the cleanup of the dissolved plume does not appear to be technically practicable in the presence of DNAPL (see Technical Impracticability [TI] Waiver, Appendix E, JGWFS [CH2M HILL, 1998] and the ROD [EPA, 1999]). In addition, the predictive portion of the run for this data gap analysis did not include contaminant sources, because it was assumed that they were contained by the remedial wellfield and did not contribute to the mass and volume of the chlorobenzene plume. The JGWFS runs included contaminant sources. While this simplification of the mass and volume estimates is acceptable for the purposes of the data gap analysis, these estimates will be revised to account for the TI waiver zone and contaminant sources at the RD model development and final wellfield design phase of this project.

In addition to the calibration targets (see Section 3, Initial Calibration), two sets of predictive targets pertaining to the mass and volume reduction were used for the "best-case" and "worst-case" runs performed during predictive calibration. For the best case, the predictive targets for the mass and volume reduction rates were consistent with the requirements of the ROD (EPA, 1999) and results obtained by Scenario 5 of the JGWFS (CH2M HILL, 1998). For the worst case, the predictive targets for mass and volume reduction rates were established to result in a 100 percent increase of the original plume mass and volume. Both sets of predictive targets were developed for the MBFC and Gage Aquifer, which are the target aquifers for the proposed remedial actions. The increase in mass and volume of the plume in the MBFC and Gage Aquifer would only be possible if pumping of the remedial wellfield would fail to contain and remove the contaminant mass, while inducing contaminant migration from the UBF into the MBFC and Gage Aquifer. The objective of the

worst case run was to assess which geologic conditions (if any) could result in such failure of the wellfield performance and determine the presence of these conditions in the field, if possible.

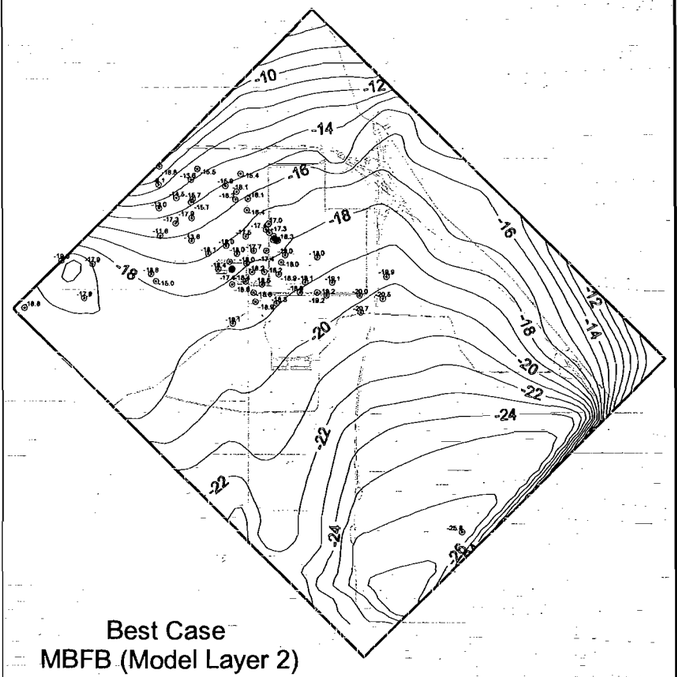
4.1.2 Calibration Results for the Best and Worst Case

Calibrated water levels and chlorobenzene concentrations for the best and worst case are compared in Figures 4-1 and 4-2, respectively. The scatter plots of measured and simulated water levels for the best and worst case are shown in Figures 4-3 and 4-4. The calibrated distribution of horizontal hydraulic conductivity for the best and worst case is shown in Figures 4-5a and 4-5b, and the calibrated distribution of vertical hydraulic conductivity is shown in Figures 4-6a and 4-6b. The calibrated transport parameters for the best and worst case are presented in Table 4-1. Based on these results, it appears that both the best- and worst-case models are reasonably well calibrated and have realistic parameters.

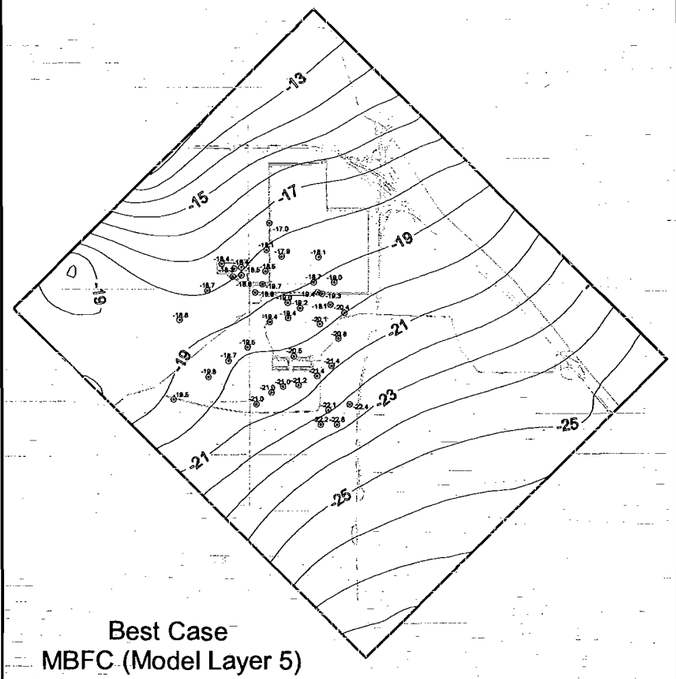
A comparison of the calibrated distributions of model parameters reveals differences in the distribution of horizontal and vertical hydraulic conductivity between the best and the worst case. A zone of lower hydraulic conductivity was assigned to the best case by PEST west of the chlorobenzene plume in the MBFC. Higher values of hydraulic conductivity in this area were assigned to the worst case. A zone of low hydraulic conductivity in this area could improve the performance of injection wells in creating a groundwater mound, which serves as a barrier to contaminant migration. In the best case, PEST also assigned a lower hydraulic conductivity to the LBF beneath the MBFC chlorobenzene plume than in the worst case. This would result in less vertical migration of contaminants from the MBFC into the Gage Aquifer, and, consequently, in better performance of the remedial wellfield.

The recharge distribution for both cases was essentially the same. The greatest differences were observed in transport parameters. Calibrated values of dispersivity were higher for the worst case in the aquifer units (MBFB and MBFC), and lower in the MBFM . Calibrated values of Kd, and consequently, R*, were higher for the worst case than those for the best case in both the MBFB and MBFC. The differences in transport parameters between the best and worst cases are consistent with the differences in predictive targets for these simulations. The higher dispersivity assigned by PEST to the worst case simulation causes the chlorobenzene plume to spread more within the aquifers, complicating containment and resulting in lower rates of volume reduction, which is consistent with the predictive target of increasing the volume of the contaminant plume. Higher Kd and R cause sorption of contaminants on soil particles of the aquifer formation, resulting in lower rates of mass removal for the worst case, which is also consistent with the predictive target for the failure of mass removal.

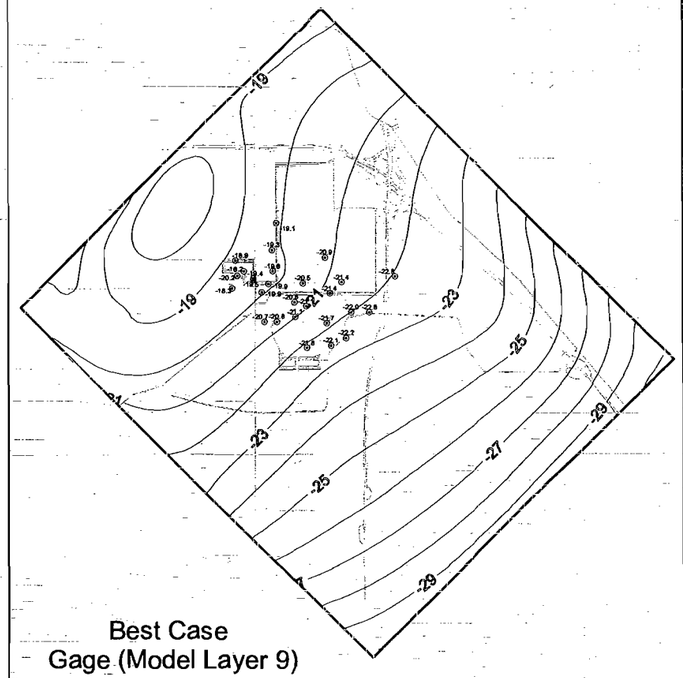
* R is calculated by MT3DMS based on Kd and bulk density values.



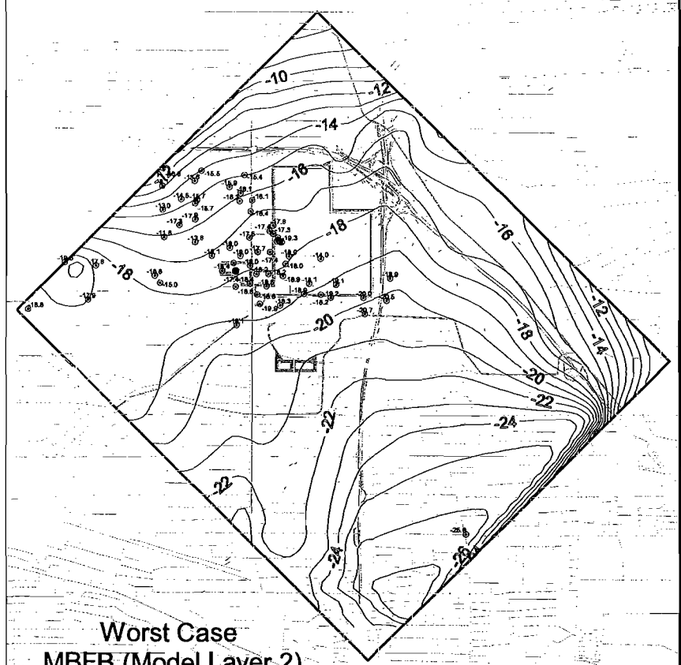
Best Case
MBFB (Model Layer 2)



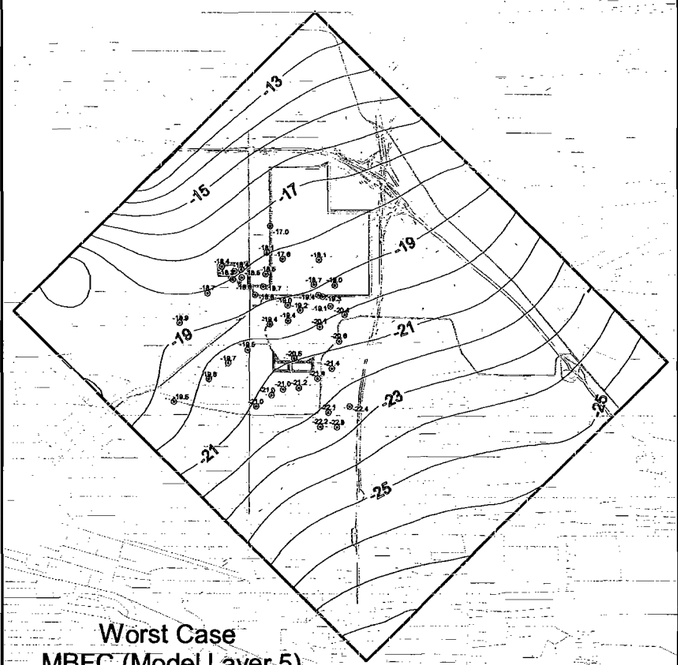
Best Case
MBFC (Model Layer 5)



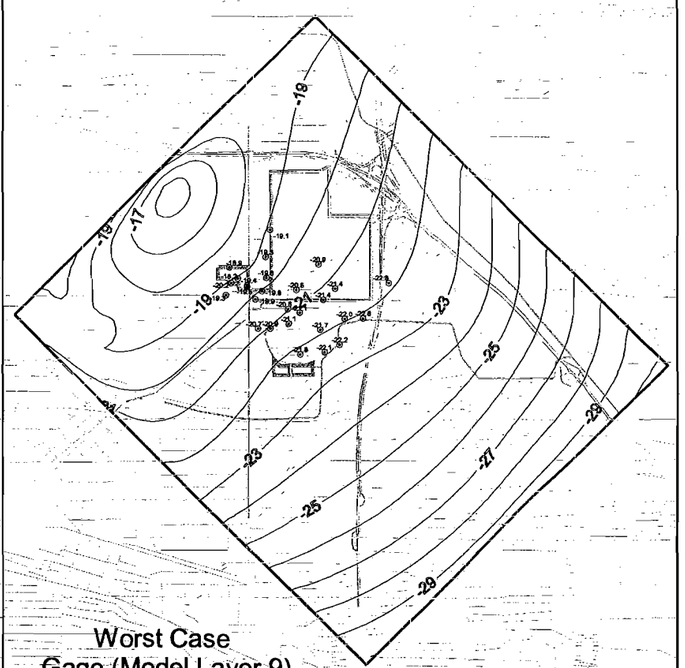
Best Case
Gage (Model Layer 9)



Worst Case
MBFB (Model Layer 2)



Worst Case
MBFC (Model Layer 5)



Worst Case
Gage (Model Layer 9)

Calibration Well Locations with Measured October 1995 Water Levels (ft msl)

Simulated October 1995 Water Level Contours (ft msl)

Streets

Site Location

Model Domain

Base Map

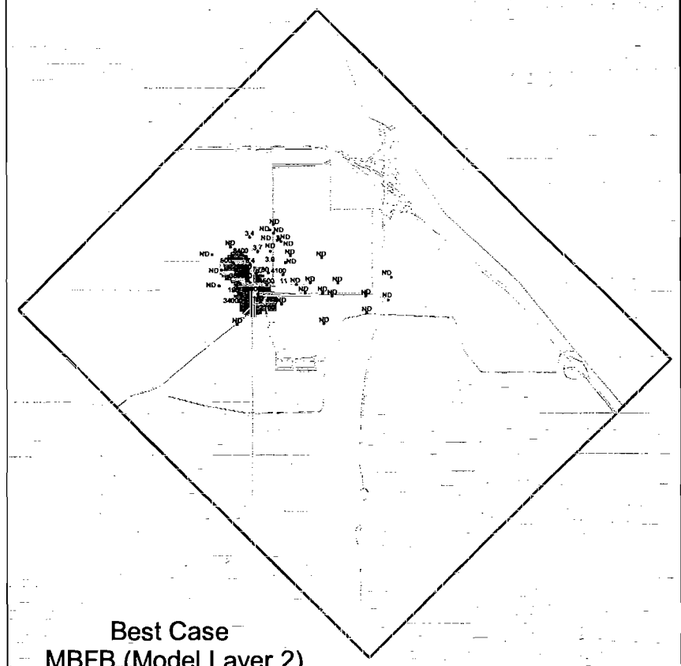
N

5000 Feet 5000 Feet

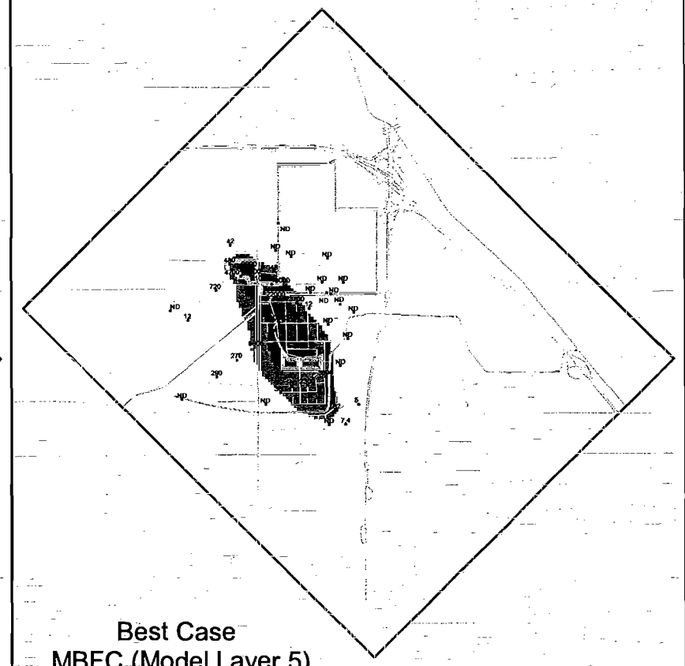
Figure 4-1
Comparison of Simulated Water Level Contours

Best and Worst Cases for Mass and Volume Reduction

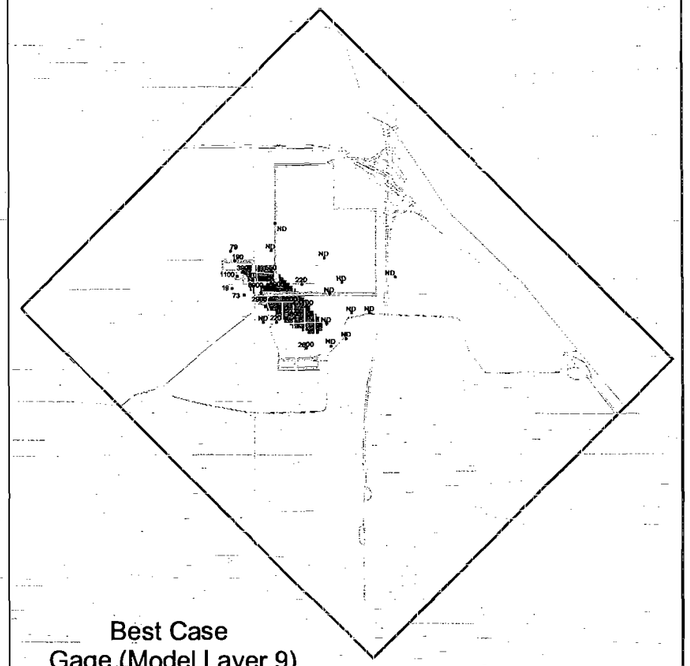
J:\Montrose_GW_Remedy\plots\092904_report_plot5



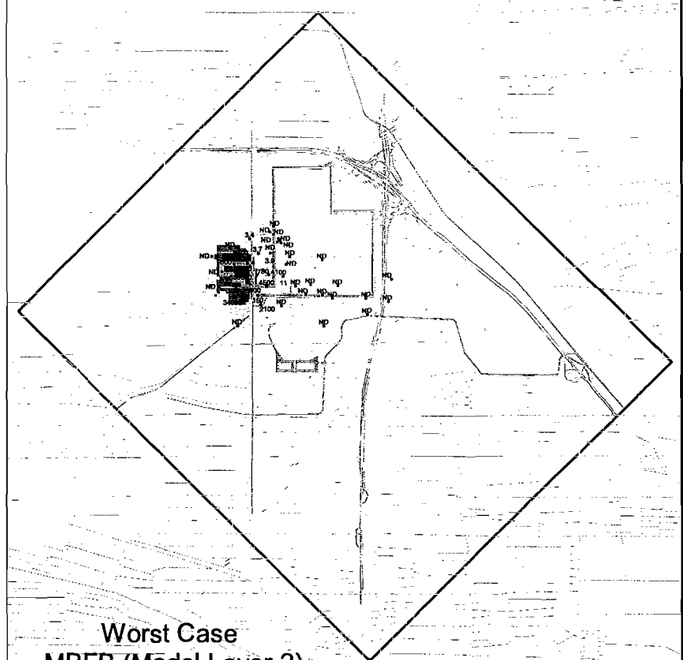
Best Case
MBFB (Model Layer 2)



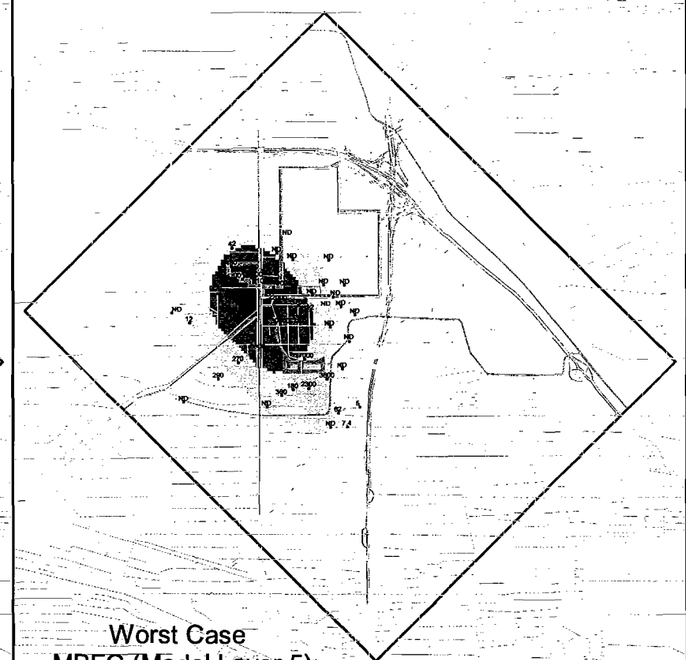
Best Case
MBFC (Model Layer 5)



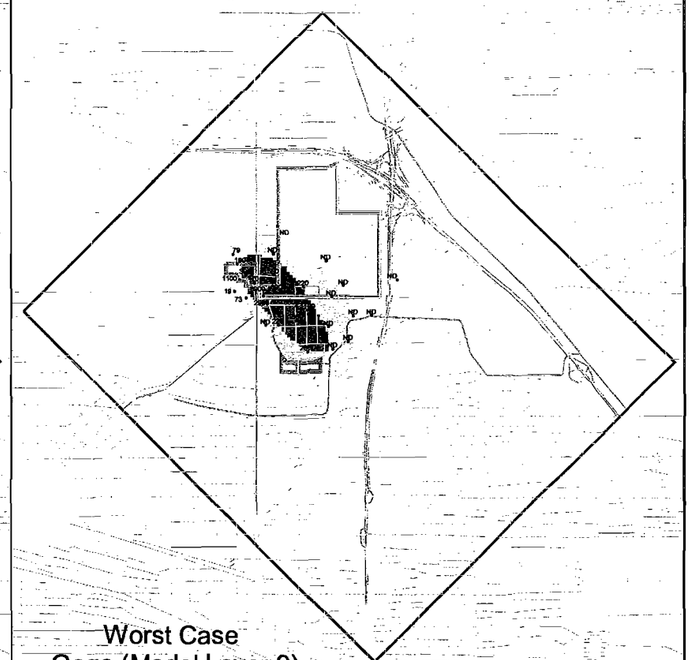
Best Case
Gage (Model Layer 9)



Worst Case
MBFB (Model Layer 2)



Worst Case
MBFC (Model Layer 5)



Worst Case
Gage (Model Layer 9)

· Observed Chlorobenzene Concentrations (µg/L)

Streets
Site Location
Model Domain
Base Map

Color Codes for Simulated Chlorobenzene Concentrations (µg/L)

70 - 1000
1000 - 5000
5000 - 10000
10000 - 100000
100000 - 1000000

N

5000 Feet 5000 Feet

Figure 4-2
Comparison of Simulated Chlorobenzene Plumes (1945 - 1995)

Best and Worst Cases for Mass and Volume Reduction

J:\Montrose_GW_Remedy\plots\092904_report_plot6

FIGURE 4-3
Simulated vs. Measured Water Levels – Best Case

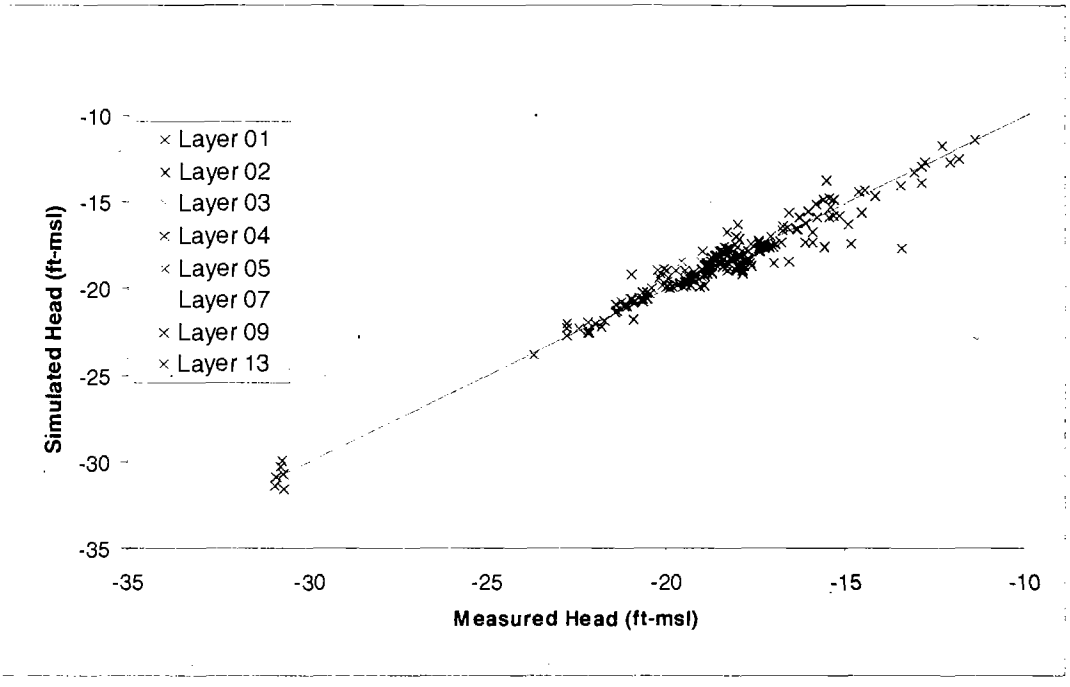
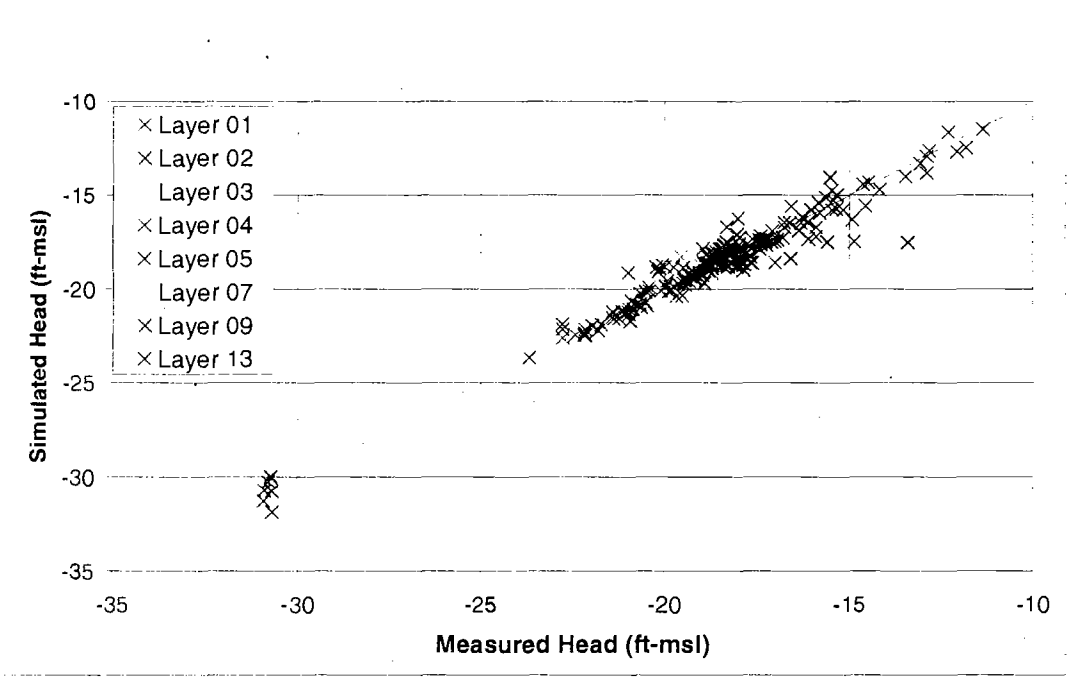
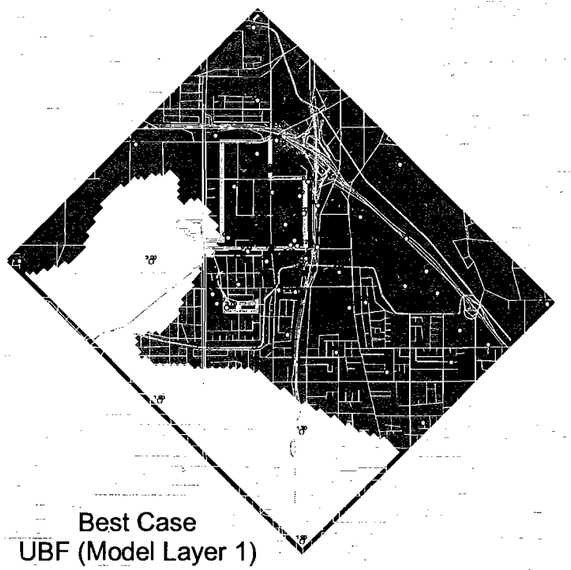
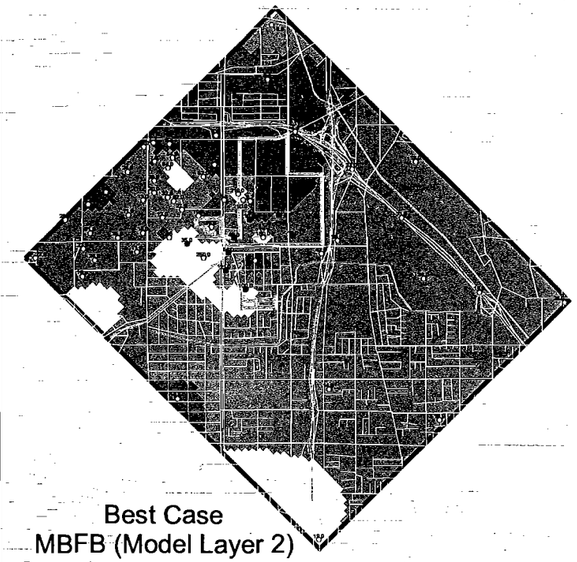


FIGURE 4-4
Simulated vs. Measured Water Levels – Worst Case

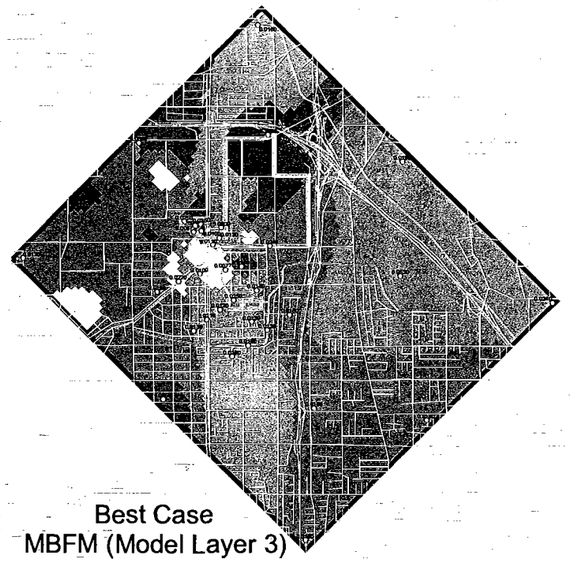




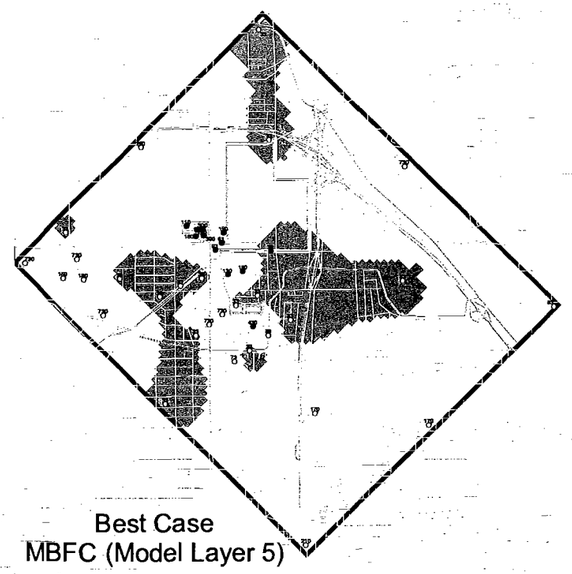
Best Case
UBF (Model Layer 1)



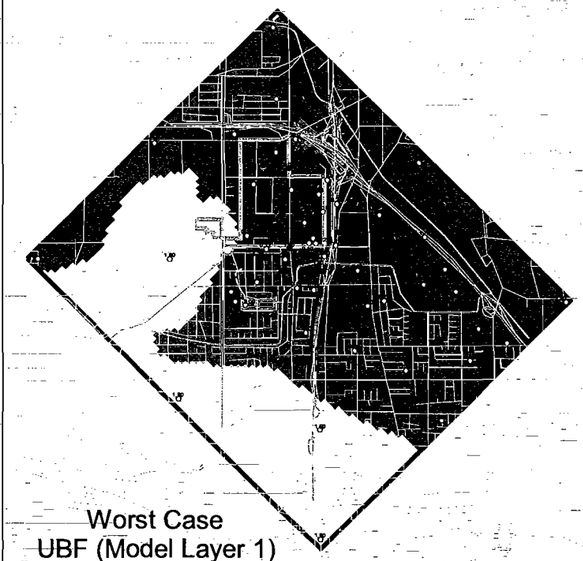
Best Case
MBFB (Model Layer 2)



Best Case
MBFM (Model Layer 3)

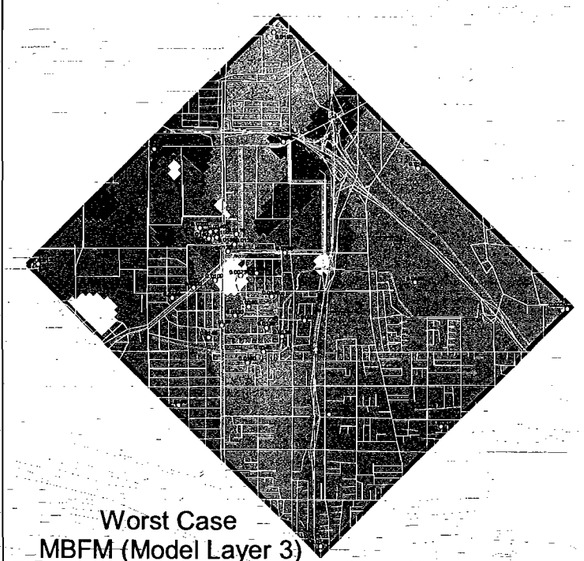


Best Case
MBFC (Model Layer 5)

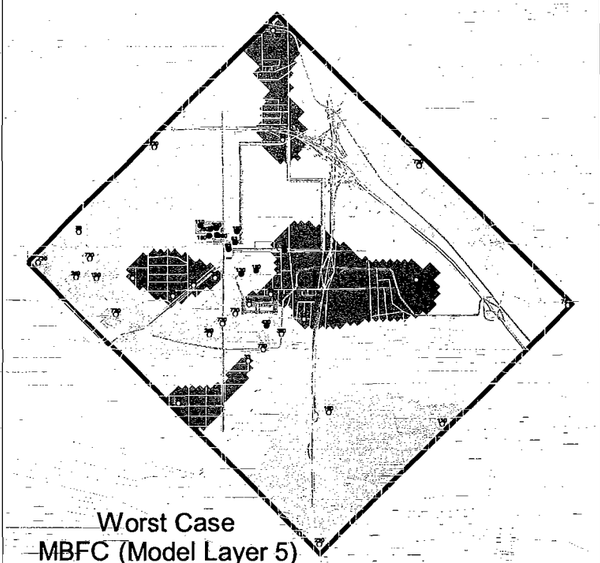


Worst Case
UBF (Model Layer 1)

Worst Case
MBFB (Model Layer 2)



Worst Case
MBFM (Model Layer 3)



Worst Case
MBFC (Model Layer 5)

- Pilot Points, Values from Model Calibration
- Pilot Points, Values from Aquifer Tests
- Streets
- Site Location
- Model Domain
- Base Map

Color Codes for Horizontal Hydraulic Conductivity (ft/day)

- 0.001 - 0.01
- 0.01 - 0.1
- 0.1 - 1
- 1 - 10
- 10 - 50
- 50 - 100
- 100 - 300
- 300 - 500
- 500 - 1000

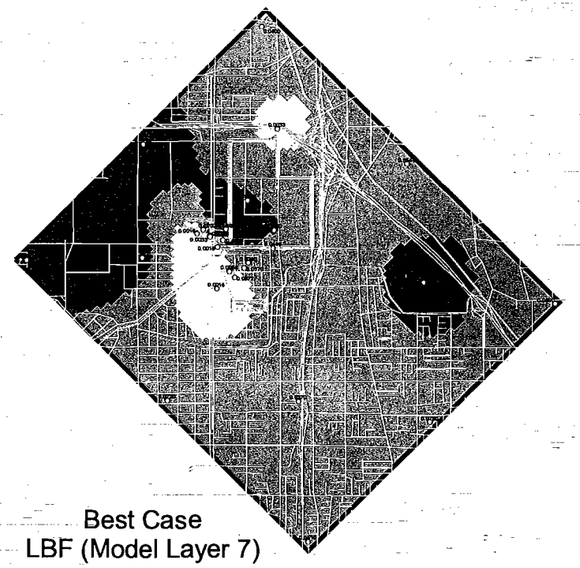
N

6000 Ft 6000 Ft

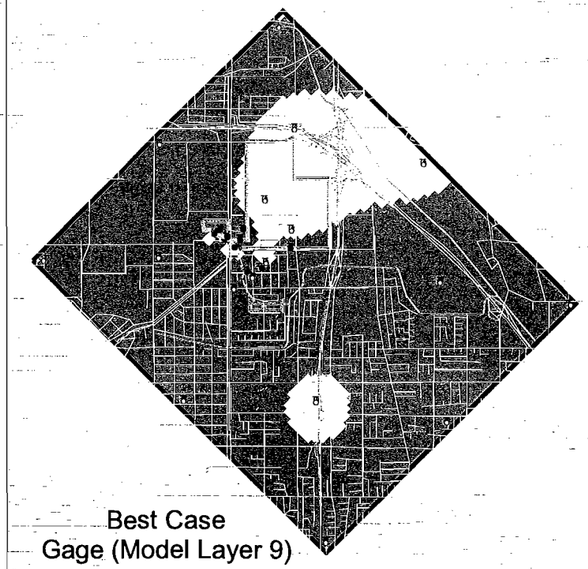
Figure 4-5a
Comparison of Calibrated Horizontal Hydraulic Conductivity

Best and Worst Cases for
Mass and Volume Reduction

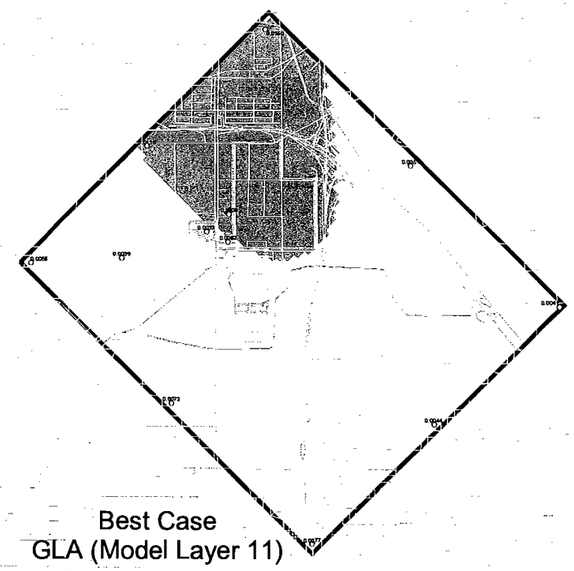
J:\Montrose_GW_Remedy\Plots\092904_report_Plot7



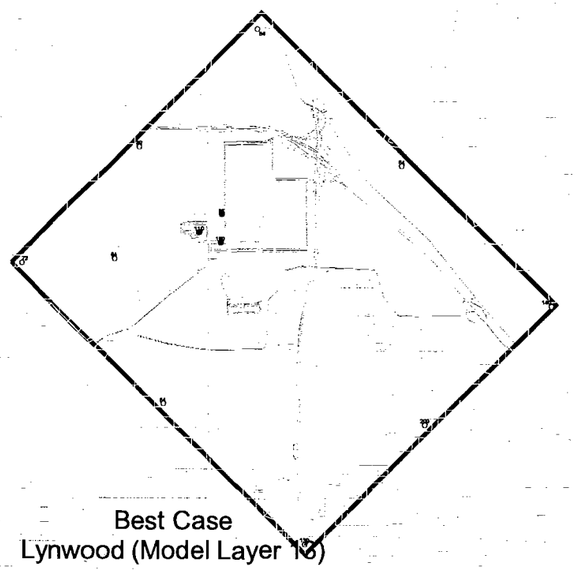
Best Case
LBF (Model Layer 7)



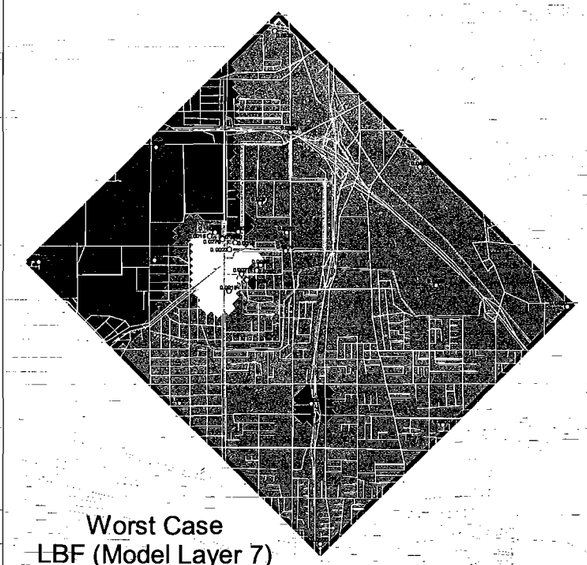
Best Case
Gage (Model Layer 9)



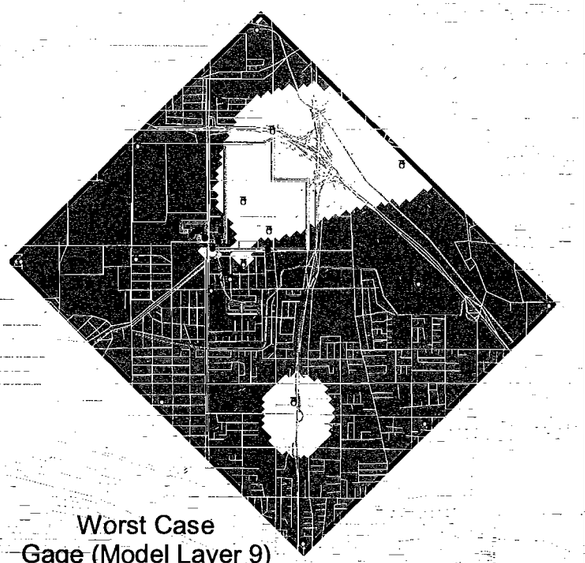
Best Case
GLA (Model Layer 11)



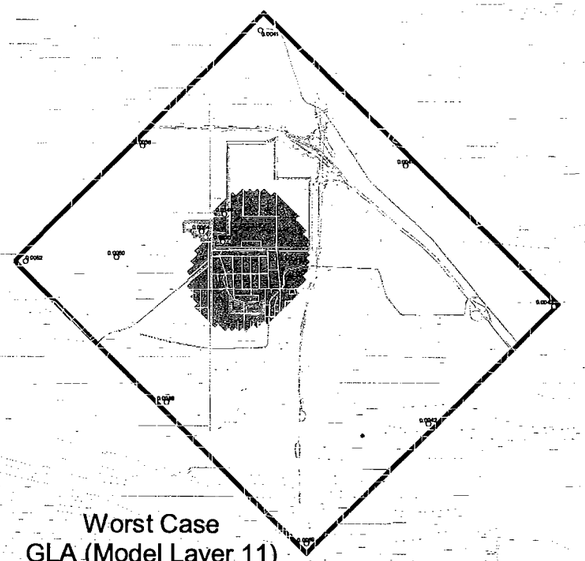
Best Case
Lynwood (Model Layer 13)



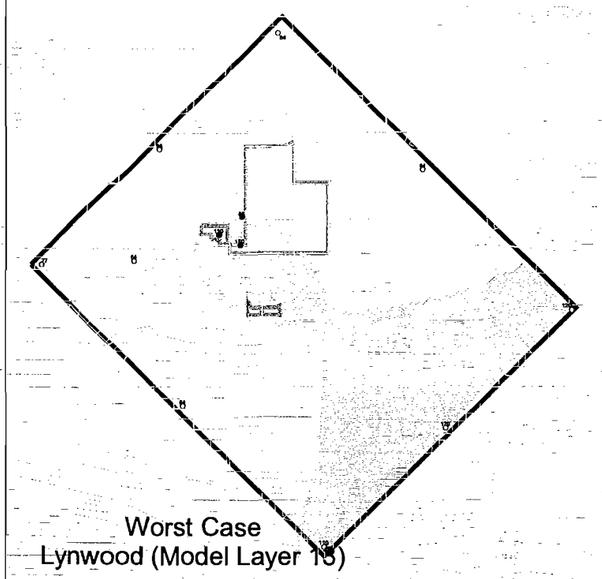
Worst Case
LBF (Model Layer 7)



Worst Case
Gage (Model Layer 9)



Worst Case
GLA (Model Layer 11)



Worst Case
Lynwood (Model Layer 13)

- ◦ Pilot Points, Values from Model Calibration
- • Pilot Points, Values from Aquifer Tests
- Streets
- Site Location
- Model Domain
- Base Map

Color Codes for Horizontal Hydraulic Conductivity (ft/day)

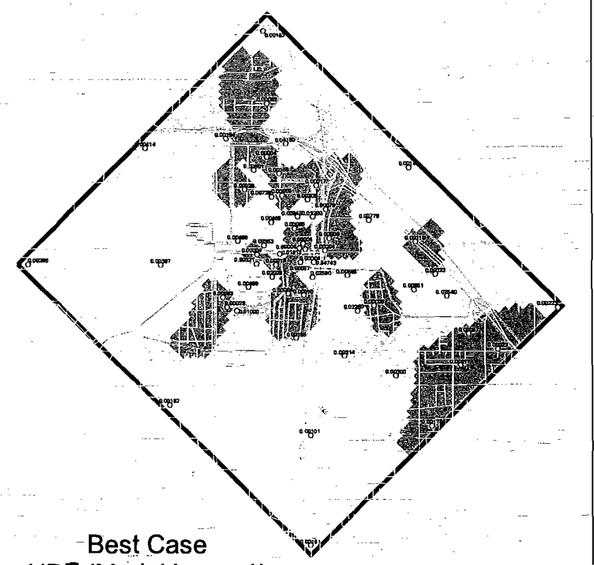
0.001 - 0.01	50 - 100
0.01 - 0.1	100 - 300
0.1 - 1	300 - 500
1 - 10	500 - 1000
10 - 50	

N

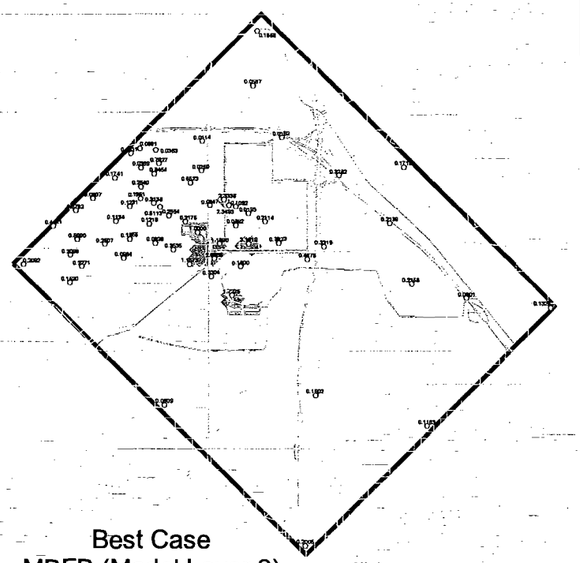
6000 Ft 6000 Ft

Figure 4-5b
Comparison of Calibrated Horizontal Hydraulic Conductivity

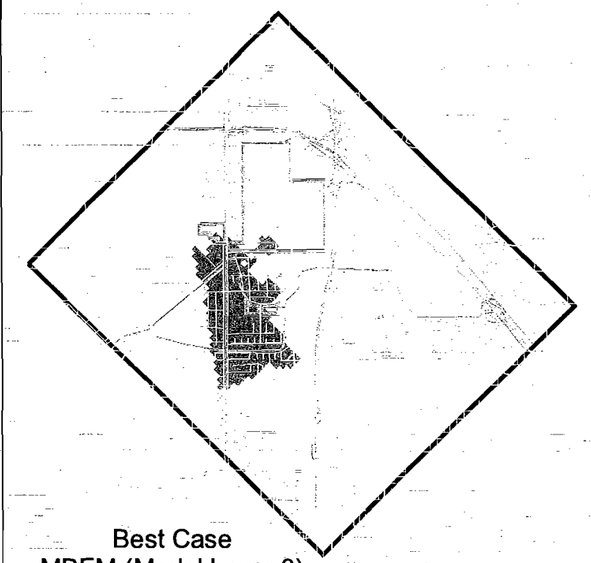
Best and Worst Cases for
Mass and Volume Reduction



Best Case
UBF (Model Layer 1)

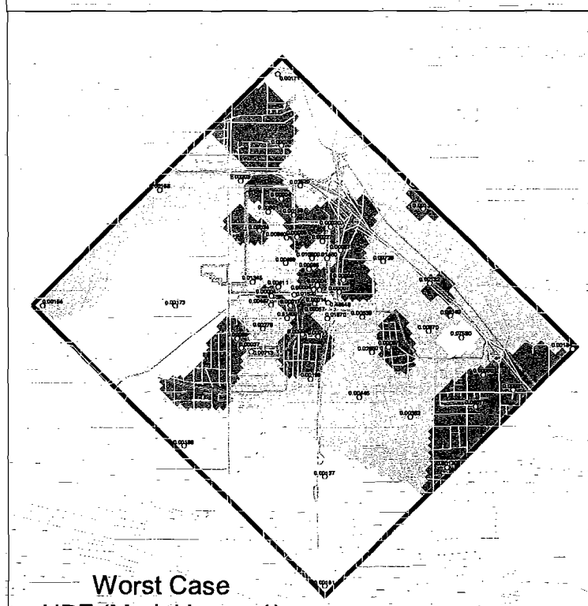


Best Case
MBFB (Model Layer 2)

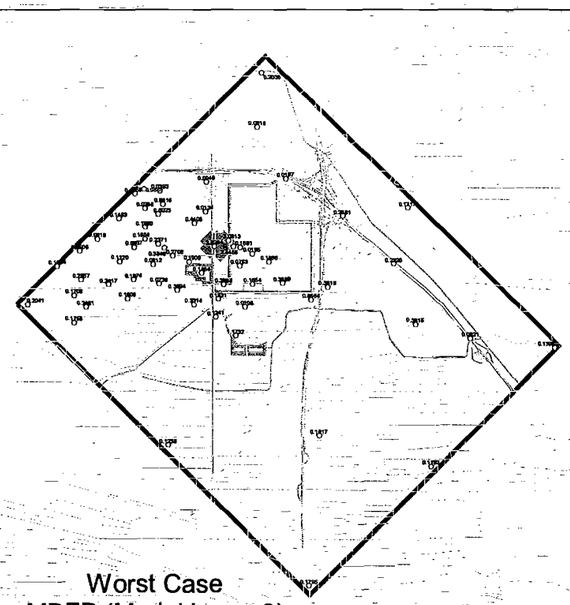


Best Case
MBFM (Model Layer 3)

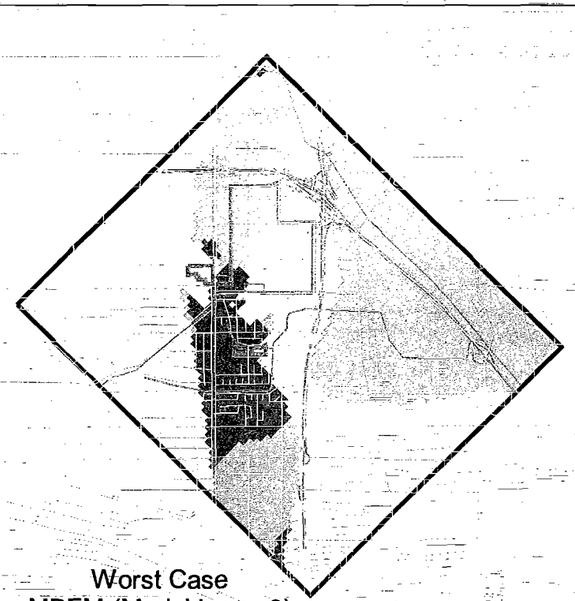
Best Case
MBFC (Model Layer 5)



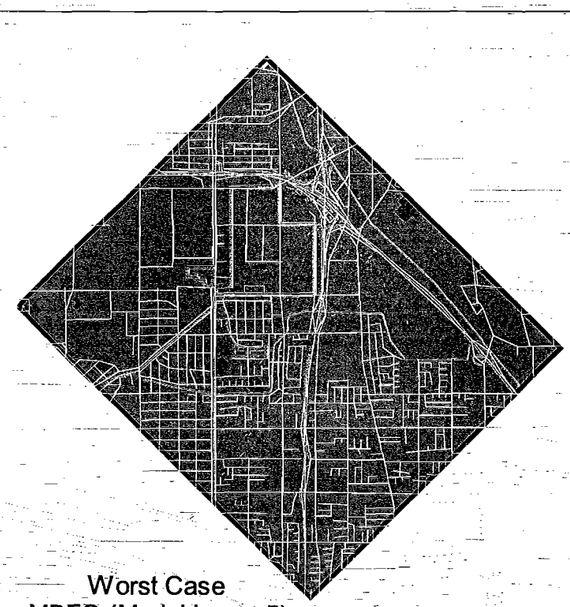
Worst Case
UBF (Model Layer 1)



Worst Case
MBFB (Model Layer 2)



Worst Case
MBFM (Model Layer 3)



Worst Case
MBFC (Model Layer 5)

- Pilot Points, Values from Model Calibration
- Pilot Points, Values from Aquifer Tests
- Streets
- Site Location
- Model Domain
- Base Map

Color Codes for Vertical Hydraulic Conductivity (ft/day)

- <0.001
- 0.001 - 0.01
- 0.01 - 0.1
- 0.1 - 1
- 1 - 10
- 10 - 50
- 50 - 100
- 100 - 300
- 300 - 500

N

6000 Ft 6000 Ft

Figure 4-6a
Comparison of Calibrated Vertical Hydraulic Conductivity

Best and Worst Cases for
Mass and Volume Reduction

TABLE 4-1
Calibrated Transport Parameters for the Best and Worst Case

Best Case

HSU	Porosity	D_L (ft)	D_T/D_L	D_Z/D_L	Chlorobenzene Source (ug/L)	Chlorobenzene Kd (mL/g)	R (n_{eff} = n)	p-CBSA Source (ug/L)	Bulk Density (g/cm3)
UBF	0.31	86	0.5	0.001		0.0053	1.03		1.49
MBFB	0.27	4.4	0.31	0.0003	100000	0.017	1.09	100000	1.49
MBFM	0.30	75			36000	0.021	1.09	100000	1.25
MBFC	0.27	6	1	0.0004	24000	0.01	1.06	100000	1.59
LBF	0.30	3.4	0.5	0.0006	1100	0.097	1.49	22000	1.52
Gage	0.10	11	0.34	0.001	730	0.099	2.51	4500	1.53
GLA	0.30	30	0.5	0.0009		0.73	4.67		1.51
Lynwood	0.30	96	0.5	0.001		0.053	1.31		1.76

Worst Case

HSU	Porosity	D_L (ft)	D_T/D_L	D_Z/D_L	Chlorobenzene Source (ug/L)	Chlorobenzene Kd (mL/g)	R (n_{eff} = n)	p-CBSA Source (ug/L)	Bulk Density (g/cm3)
UBF	0.30	110	0.5	0.001		0.0053	1.03		1.49
MBFB	0.31	1000	0.19	0.0029	100000	2.1	11.09	100000	1.49
MBFM	0.30	9.8			39000	0.021	1.09	100000	1.25
MBFC	0.30	190	1	0.00014	21000	0.23	2.22	100000	1.59
LBF	0.30	20	0.5	0.00047	1100	0.082	1.42	24000	1.52
Gage	0.10	7.2	0.28	0.001	730	0.027	1.41	4300	1.53
GLA	0.30	22	0.5	0.001		0.68	4.42		1.51
Lynwood	0.30	110	0.48	0.001		0.052	1.31		1.76

4.1.3 Predictions of Mass and Volume Reduction for the Best and Worst Cases

Table 4-2 presents the comparison of the simulated percent reduction of the chlorobenzene plume volume and mass for the best- and worst-case simulations in the MBFC and Gage aquifers. It also provides the percent reduction estimates for the baseline calibration and for JGWFS Scenario 5. Figure 4-7 compares the volume retained in the aquifers after 10, 25, and 50 years for the best and worst case and the baseline calibration. Figure 4-8 compares the percentage of mass retained in the aquifers for the same scenarios.

TABLE 4-2
Comparison of the Simulated Percent Reduction of the Chlorobenzene Plume Volume and Mass

Simulation	MBFC			Gage		
Percent Volume Reduction						
	10 Years	**25 Years**	**50 Years**	**10 Years**	**25 Years**	**50 Years**
JGWFS	35%	68%	94%	44%	68%	81%
Best	75%	97%	99%	100%	100%	100%
Baseline	39%	66%	83%	100%	100%	100%
Worst	5%	31%	42%	97%	100%	100%
Percent Mass Reduction						
	10 Years	**25 Years**	**50 Years**	**10 Years**	**25 Years**	**50 Years**
JGWFS	76%	89%	92%	68%	85%	86%
Best	94%	99%	>99%	100%	100%	100%
Baseline	68%	86%	92%	100%	100%	100%
Worst	57%	74%	81%	99%	100%	100%

FIGURE 4-7
Comparison of the Percentage of Volume Retained in the Aquifer

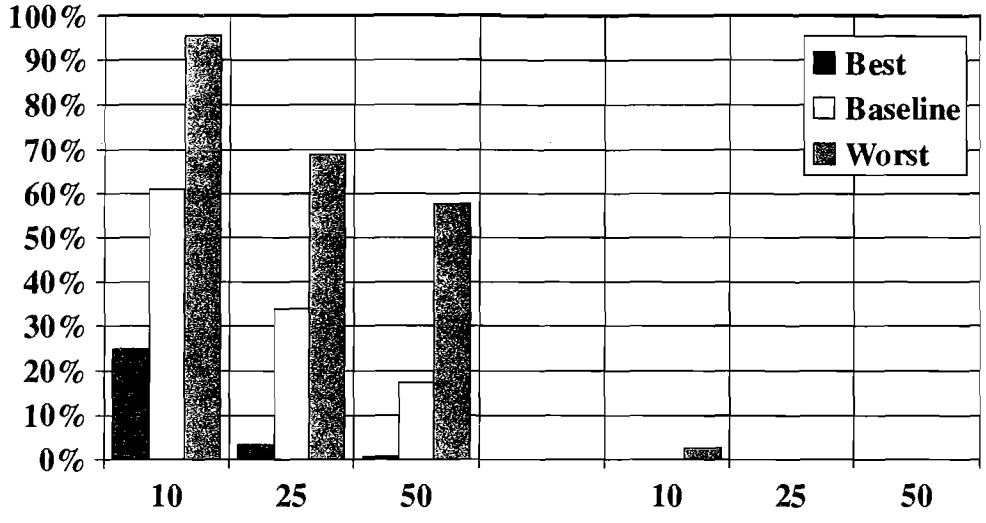
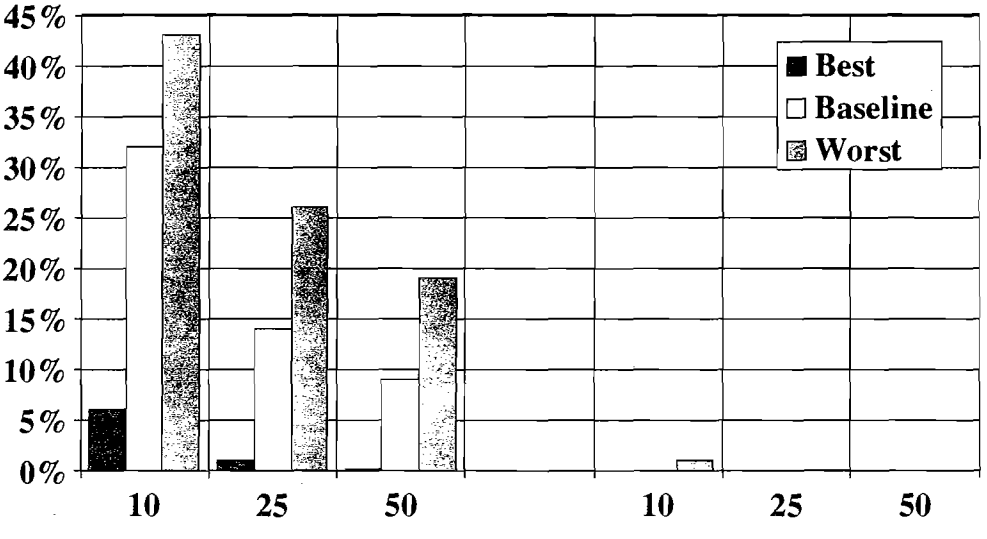


FIGURE 4-8
Comparison of the Percentage of Mass Retained in the Aquifer



Figures 4-9 and 4-10 present simulated distributions of chlorobenzene after 10, 25, and 50 years of remedial pumping using the best- and the worst-case models. Note that these figures include concentrations that are below 70 µg/L (i.e., below MCL). These concentrations are shown to facilitate the comparison between the best- and worst-case results. However, as discussed above, below-MCL concentrations were not considered in the estimates of mass and volume reduction rates.

The results of the predictive simulations indicate a significant difference between the performance of the remedial wellfield with regard to mass and volume reduction for the best and worst case in the MBFC. While the best-case prediction meets and exceeds the ROD requirements, the worst-case prediction results in the significantly slower rates of mass and volume removal in this unit. These results indicate that there is a considerable uncertainty associated with the ability of the model to predict mass and volume removal rates for the chlorobenzene plume in the MBFC, because three equally calibrated models (baseline, best case, and worst case), which are based on equally viable distributions of model parameters, produce vastly different results with regard to this prediction. This uncertainty can only be reduced if additional site-specific data could be obtained for the parameters, which have the highest contribution to this uncertainty. These parameters are discussed below.

The results of the predictive simulations of mass and volume removal in the Gage Aquifer, on the other hand, are similar for the baseline calibration, and the best and worst cases. These simulations indicate that the certainty of the plume cleanup to below-MCL levels in the Gage Aquifer after only 10 years is fairly high in the absence of contaminant sources. The JGWFS model estimated lower rates of mass and volume removal in the Gage Aquifer, because it accounted for the contaminant sources. As discussed above, the contaminant sources and the TI waiver zone will be considered in the future simulations of cleanup rates for the design of the remedial wellfield.

As mentioned above, a 70-µg/L contour of chlorobenzene concentrations was defined based on the available 1990 through 1995 data, which were also used for the JGWFS modeling (more recent 2004 baseline data were not available at the time of these analyses). In the Gage Aquifer, this contour was "closed" south of well G-19 for the purposes of these analyses, although the actual extent of the plume south of G-19 needs to be delineated as part of the proposed data acquisition.

4.1.4 Model Uncertainty for Predictions of Mass and Volume Reduction

As discussed in Section 4.1.2, differences in model parameters for the best and worst cases were compared qualitatively to identify parameters that change the most between these two cases and, therefore, have the most effect on the predictive outcome. The biggest differences were observed in values of dispersivity and Kd.

In addition to the qualitative comparison of model parameters, the sensitivity of predictive targets to each model parameter was calculated as part of the predictive calibration process. The estimated sensitivities were then used to assess the contribution of each parameter to the predictive uncertainty of the model. As discussed in Section 2.2, this assessment also accounted for the variability of each parameter in the subsurface. A detailed discussion of the methodology used to estimate the contribution of different parameters to the predictive

uncertainty of the model is presented in Appendix A and discussed in Moore and Doherty (2004). The estimated contributions of model parameters to the predictive uncertainty pertaining to the rate of volume reduction are presented in Table 4-3.

TABLE 4-3
Contributions of Model Parameters to Predictive Uncertainty Pertaining to Volume Reduction

	Contribution to Uncertainty (i.e., predictive error variance) for Volume Reduction Targets (percent)^2	
Model Parameter	**MBFC**	**Gage Aquifer**
Total	95	79
Porosity	0.074	0.054
Dispersivity	15	3.8
Chlorobenzene source concentration	0.48	0.10
Chlorobenzene adsorption constant Kd	24	30
p-CBSA source concentration	2.0	4.61
Riverbed conductance	0.00007	0.00002
GHB heads	1.6	2.5
Recharge	0.012	0.0017
K_h in UBF	0.57	0.060
K_h in MBFB	1.6	0.15
K_h in MBFM	0.078	0.0025
K_h in MBFC	16	2.5
K_h in LBF	2.5	19
K_h in Gage Aquifer	0.82	5.6
K_h in GLA	0.36	1.3
K_h in Lynwood Aquifer	0.048	0.076
K_v in UBF	0.52	0.081
K_v in MBFB	0.39	0.013

NOTE: The contribution to uncertainty is expressed in units of variance for percent volume reduction – (percent)^2. Larger numbers indicate a greater contribution.

The contribution of model parameters to the uncertainty pertaining to the rate of mass reduction was generally consistent with that for volume reduction.

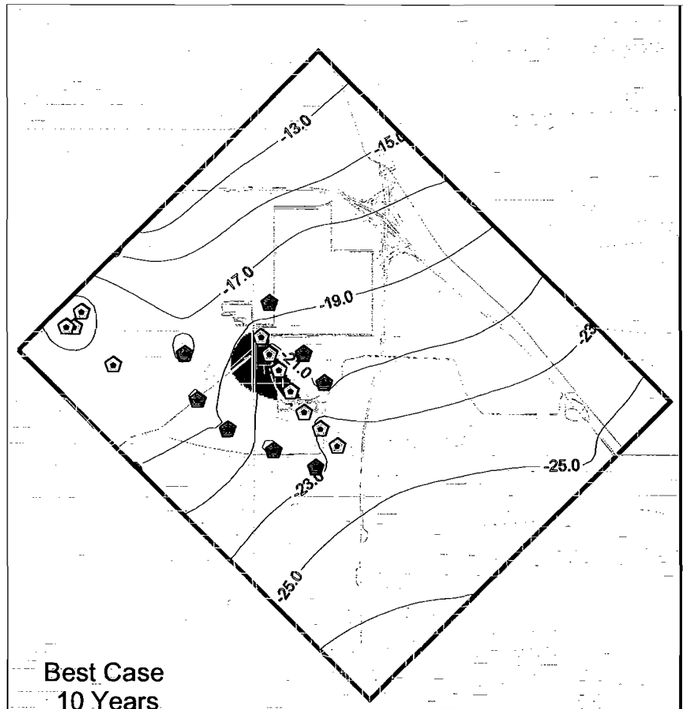
Based on these results, Kd has the highest contribution to the predictive uncertainty of the model with respect to mass and volume reduction rates in both the MBFC and Gage Aquifer. However, as discussed above, the predictive uncertainty pertaining to mass and volume reduction in the Gage Aquifer is significantly lower than that for the MBFC.

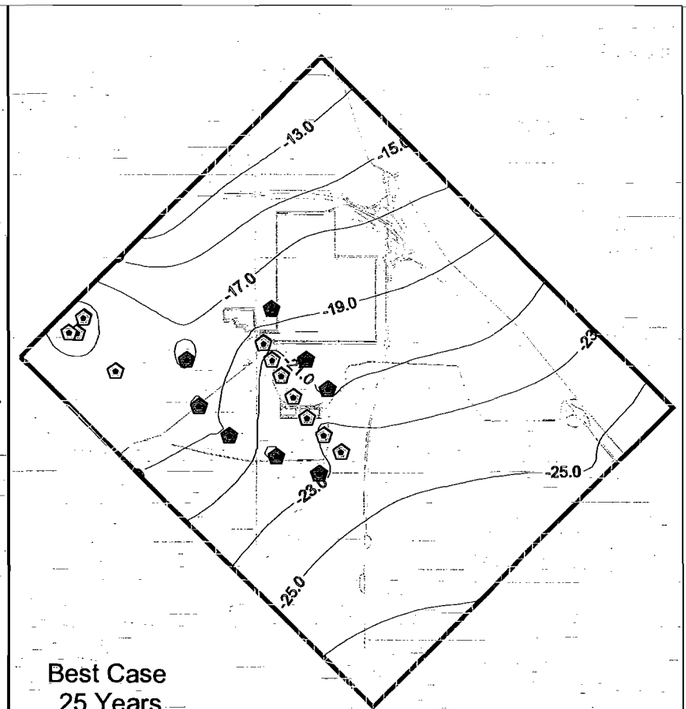
Because of the significant variability of Kd in the natural systems such as the Site, and because the field experiments required to quantify this parameter are complicated, costly, time-consuming, and ordinarily ineffective, uncertainty associated with Kd cannot be appreciably reduced. In addition, the phenomenon of "slow desorption or irreversible sorption, "** which may have a significant impact on the actual cleanup times, cannot be accounted for in the model even if field Kd data are available, because of the limitations of the MT3DMS code, which does not allow for this level of complexity in representation of dissolved/sorbed contaminant interaction. Consequently, it does not appear that predictive uncertainty with regard to chlorobenzene mass and volume reduction rates could be reduced by collecting additional field Kd data.

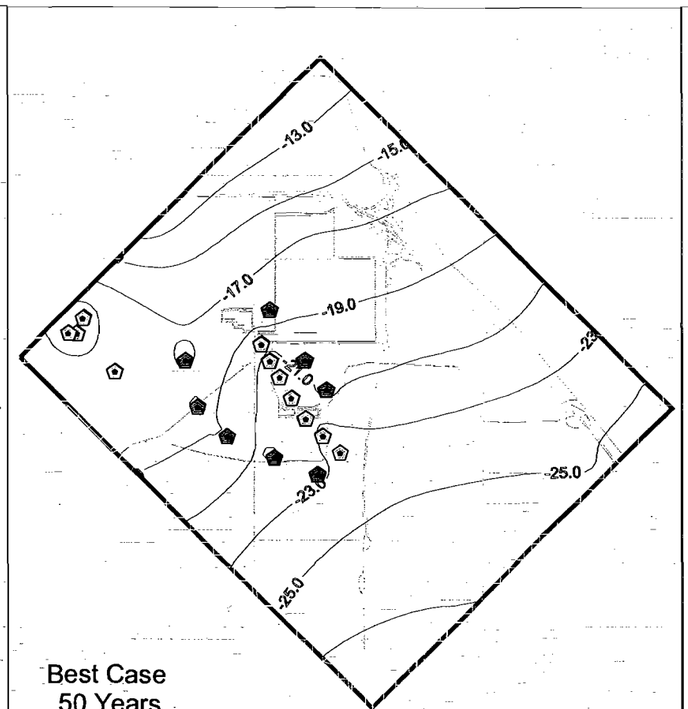
Hydraulic conductivity and dispersivity in the MBFC also contribute to the predictive uncertainty pertaining to volume reduction in the MBFC, but the contribution of these parameters is lower than that of Kd. Hydraulic conductivity in the LBF contributes to the predictive uncertainty pertaining to volume reduction in the Gage Aquifer, but this uncertainty is fairly low. While the values of dispersivity are very difficult to obtain in the field, additional hydraulic conductivity data in the MBFC can be obtained during pilot testing. Based on its contribution to predictive uncertainty, additional hydraulic conductivity data in the MBFC may reduce the predictive uncertainty of the model pertaining to volume reduction rates in the MBFC to some degree. However, as discussed above, the significant uncertainty associated with the model predictions of these performance criteria in the MBFC will still be caused by uncertainties in Kd. The need for additional field values of hydraulic conductivity in the MBFC is also discussed in Section 4.2 with regard to reducing uncertainty of containment in the MBFC.

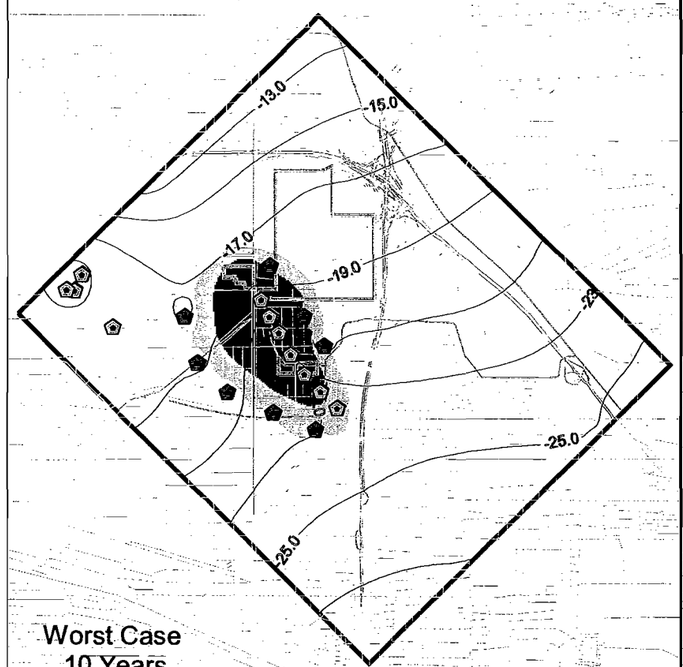
In summary, significant uncertainty is associated with the performance criteria pertaining to the rates of cleanup (i.e., mass and volume reduction) in the MBFC. Because the major contributor to this uncertainty is Kd, which cannot be easily obtained through field measurements, this uncertainty is considered to be mostly irreducible and should be accounted for by the RD and performance monitoring program. The predictive uncertainty of the model pertaining to the rates of cleanup in the Gage Aquifer is relatively low.

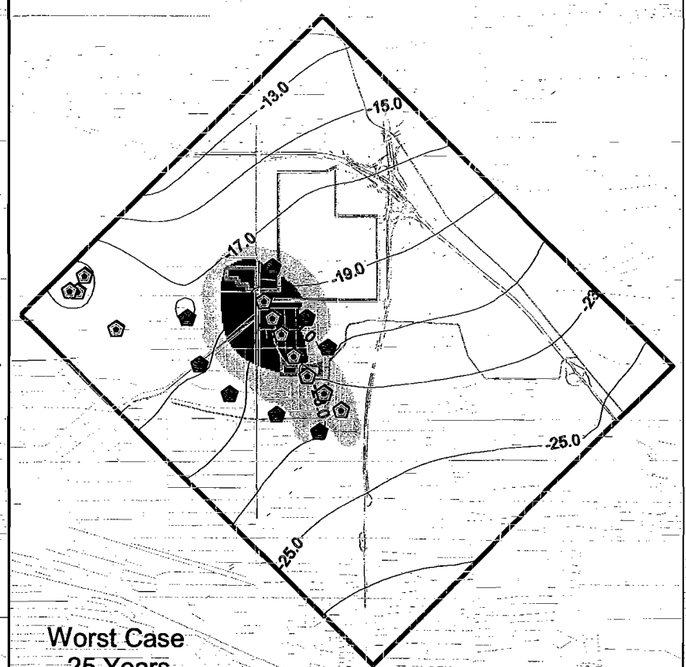
** Recent research on the ability of chemical compounds to completely desorb from a solid indicates that solid-phase contaminant concentrations can exceed the concentration predicted based on the aqueous-phase contaminant concentration and distribution coefficient (Fu et al. 1994; Kan et al., 1994; Kan et al., 1997; Pignatello and Xing, 1995). This phenomenon could be explained as slow desorption or irreversible sorption. It is reported that this situation generally happens in materials that have been in contact with contaminants for long time periods and have low solid-phase contaminant concentrations, which normally are less than 20 milligrams per kilogram (Bedient et al., 1999).

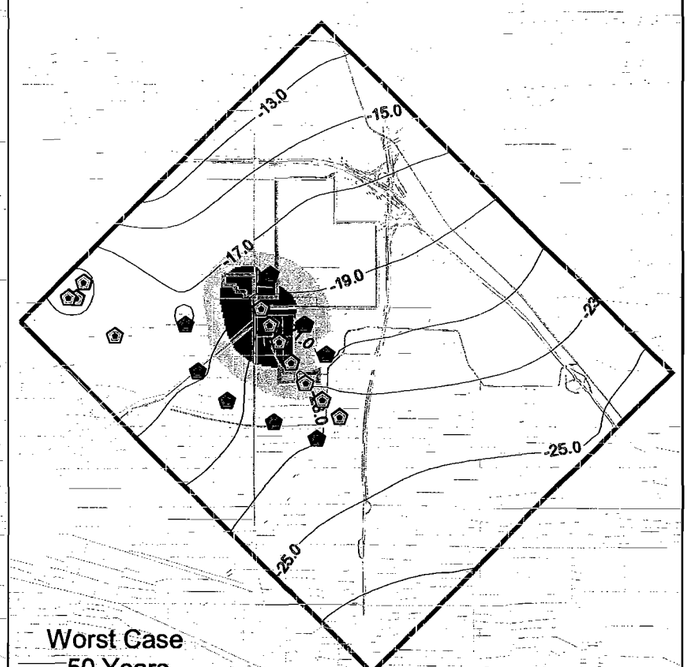












Injection Wells
Extraction Wells
Simulated Water Level Contours (ft msl)
Streets
Site Location
Model Domain
Base Map

Color Codes for Simulated Chlorobenzene Plumes (µg/L)
- <10
- 10 - 50
- 50 - 100
- 100 - 500
- 500 - 1000
- >1000

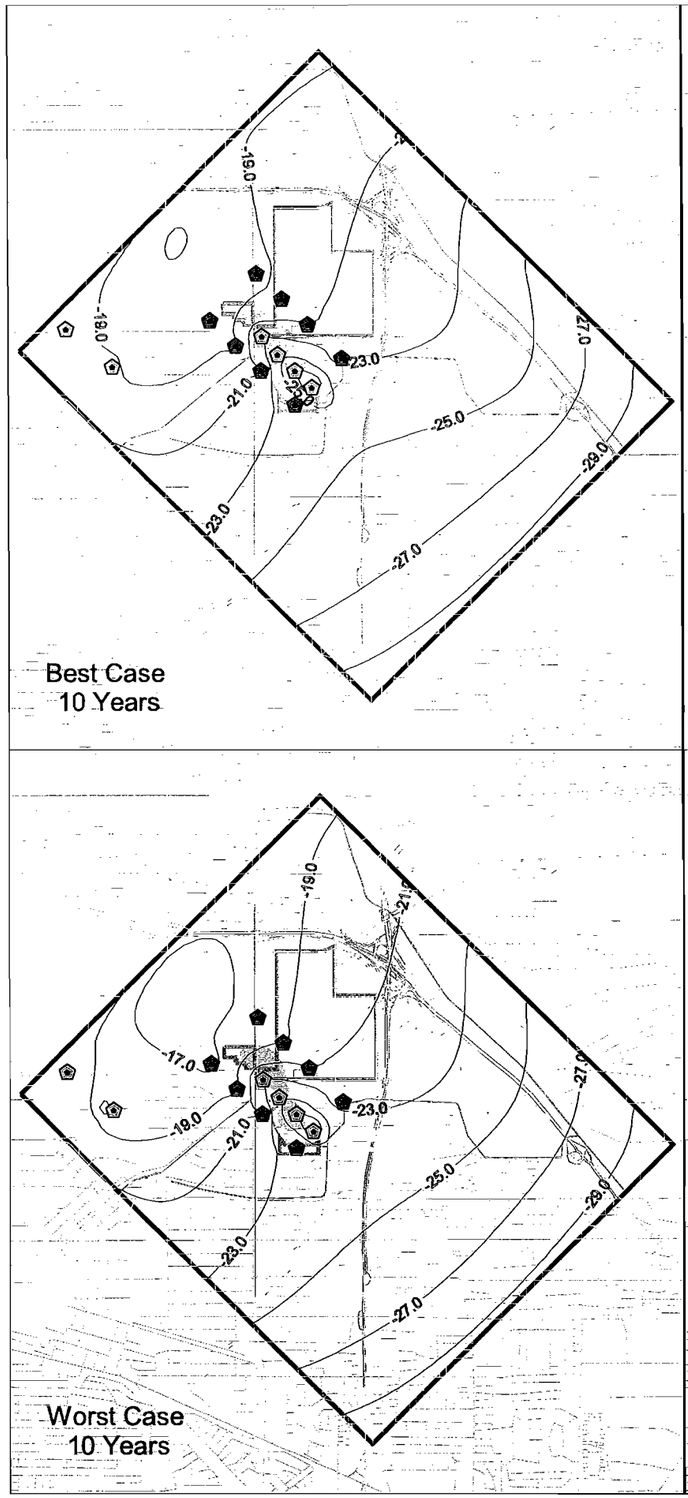
N

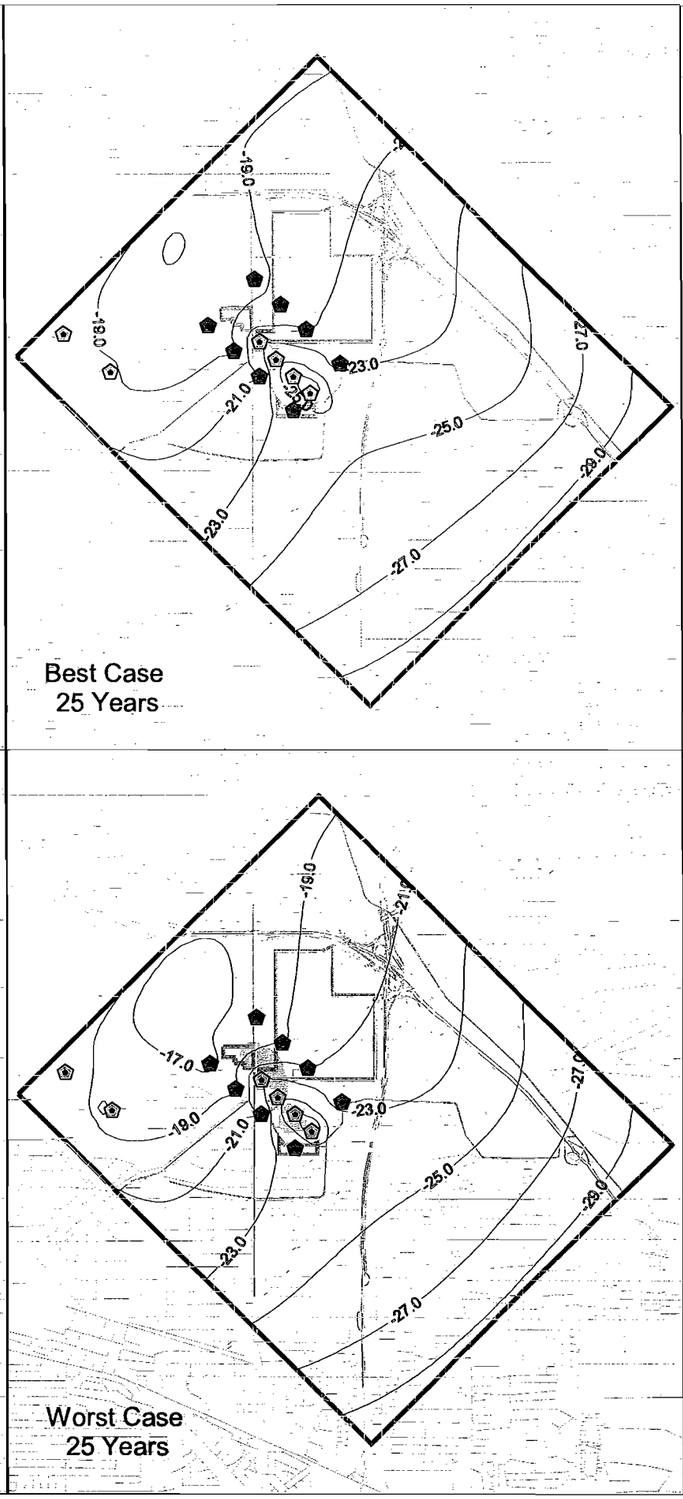
5000 Feet 5000 Feet

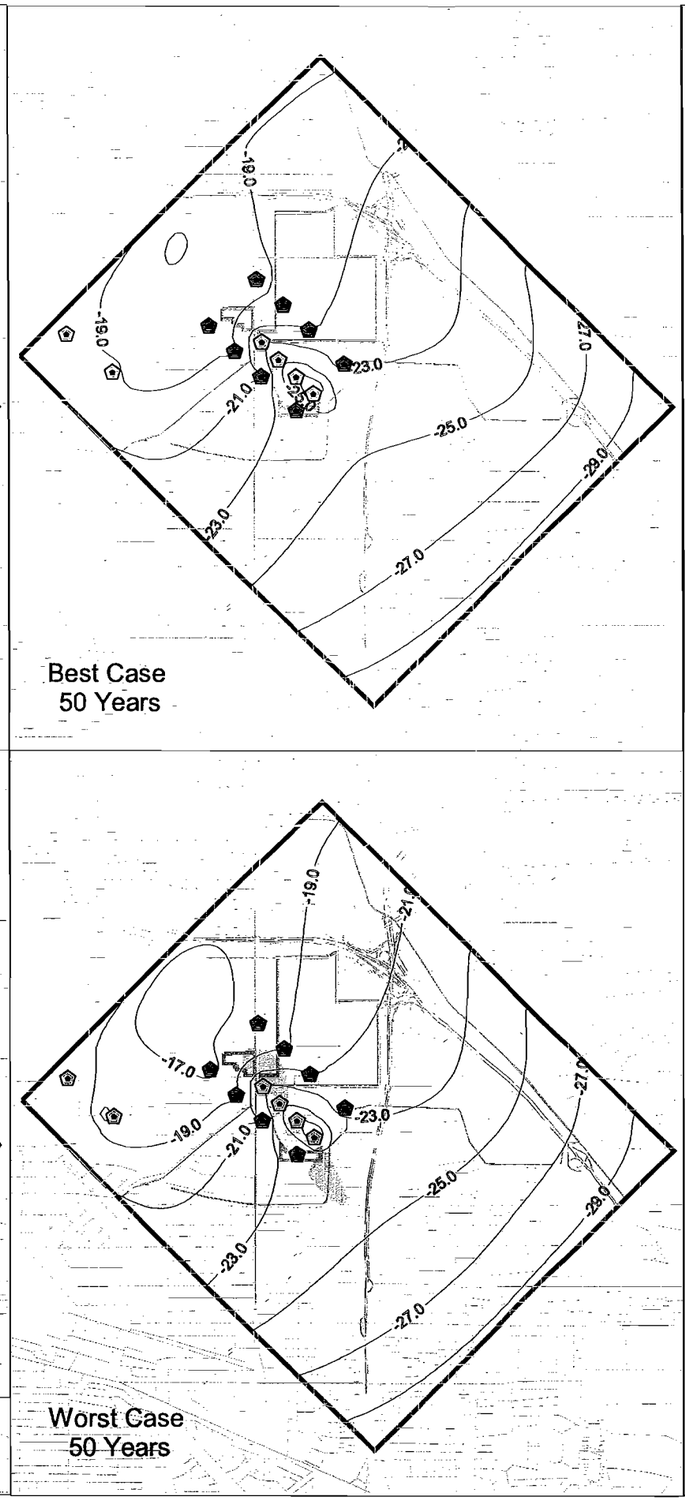
Figure 4-9
Simulated Distributions of Chlorobenzene in the MBFC after 10, 25, and 50 Years of Remedial Pumping

Best and Worst Cases for Mass and Volume Reduction

J:\Montrose_GW_Remedy\plots\092904_report_plot10







Injection Wells
Extraction Wells
Model Simulated Water Level Contours (ft msl)
Streets
Site Location
Model Domain
Base Map

Color Codes for Simulated Chlorobenzene plumes (µg/L)
- <10
- 10 - 50
- 50 - 100
- 100 - 500
- 500 - 1000
- >1000

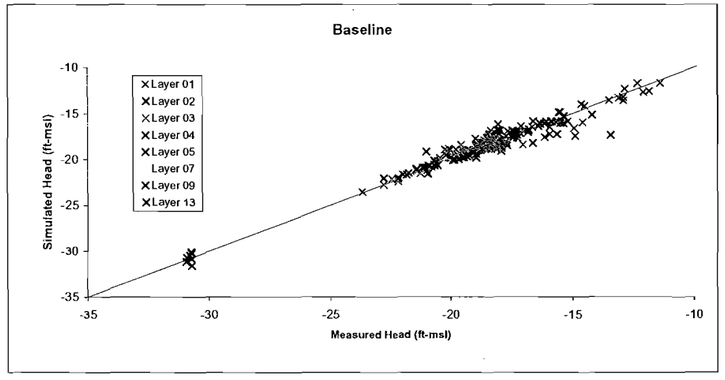
N

5000 Feet 5000 Feet

Figure 4-10
Simulated Distributions of Chlorobenzene in the Gage after 10, 25, and 50 Years of Remedial Pumping

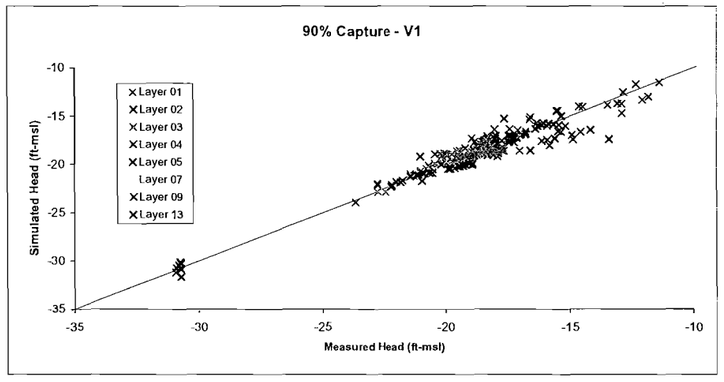
Best and Worst Cases for Mass and Volume Reduction

J:\Montrose_GW_Remedy\plots\092904_report_plot11



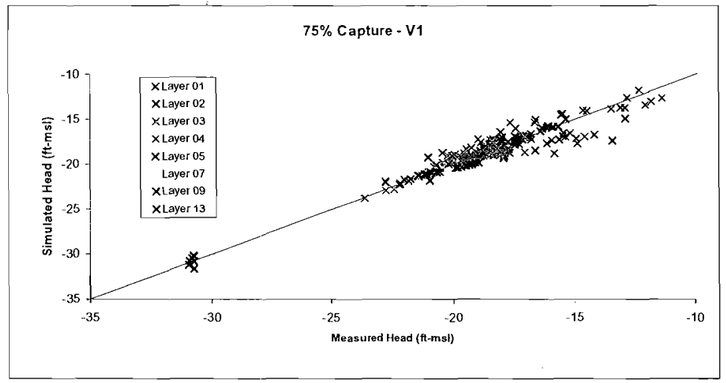
Summary Table - Baseline

Observation Group	Name	ALL	Layer01	Layer02	Layer03	Layer04	Layer05	Layer07	Layer09	Layer13
	Count	215	53	58	8	25	34	1	27	9
Observation Summary	Max	-11.4	-11.4	-13.4	-17.8	-17.0	-17.9	-20.4	-18.9	-20.9
	Min	-30.9	-23.7	-20.7	-20.0	-22.8	-22.2	-20.4	-22.8	-30.9
	Range	19.56	12.30	7.26	2.20	5.73	4.29	0.00	3.84	10.00
Residual Statistics	Max	3.95	2.57	3.95	0.45	1.40	0.64	0.21	0.50	0.90
	Min	-1.83	-1.82	-1.22	-1.12	-1.27	-1.15	0.21	-0.70	-1.83
	Mean[1]	-0.01	0.10	-0.05	-0.10	0.02	-0.05	0.21	-0.03	-0.18
	A-Mean[2]	0.45	0.59	0.47	0.36	0.59	0.23	0.21	0.29	0.60
	RMS[3]	0.66	0.80	0.77	0.48	0.71	0.34	0.21	0.34	0.78
	%RMS[4]	3.4%	6.5%	10.5%	21.8%	12.3%	7.9%		8.9%	7.8%



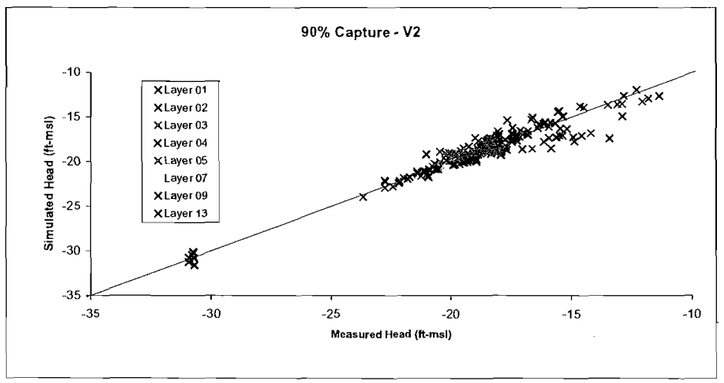
Summary Table - 90% Capture - V1

Observation Group	Name	ALL	Layer01	Layer02	Layer03	Layer04	Layer05	Layer07	Layer09	Layer13
	Count	215	53	58	8	25	34	1	27	9
Observation Summary	Max	-11.4	-11.4	-13.4	-17.8	-17.0	-17.9	-20.4	-18.9	-20.9
	Min	-30.9	-23.7	-20.7	-20.0	-22.8	-22.2	-20.4	-22.8	-30.9
	Range	19.56	12.30	7.26	2.20	5.73	4.29	0.00	3.84	10.00
Residual Statistics	Max	4.00	2.56	4.00	0.76	1.58	0.85	0.17	0.97	0.97
	Min	-2.34	-2.34	-1.09	-0.90	-1.31	-0.95	0.17	-0.78	-1.81
	Mean[1]	0.04	0.07	-0.01	0.00	0.10	-0.05	0.17	0.24	-0.13
	A-Mean[2]	0.57	0.96	0.47	0.40	0.64	0.27	0.17	0.41	0.61
	RMS[3]	0.81	1.18	0.77	0.49	0.76	0.36	0.17	0.52	0.79
	%RMS[4]	4.2%	9.6%	10.6%	22.3%	13.2%	8.5%		13.5%	7.9%



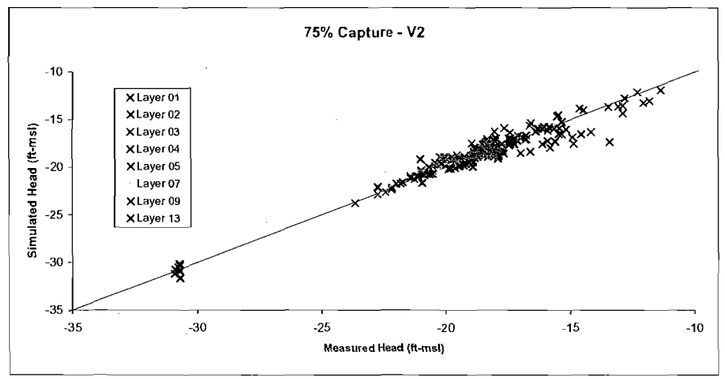
Summary Table - 75% Capture - V1

Observation Group	Name	ALL	Layer01	Layer02	Layer03	Layer04	Layer05	Layer07	Layer09	Layer13
	Count	215	53	58	8	25	34	1	27	9
Observation Summary	Max	-11.4	-11.4	-13.4	-17.8	-17.0	-17.9	-20.4	-18.9	-20.9
	Min	-30.9	-23.7	-20.7	-20.0	-22.8	-22.2	-20.4	-22.8	-30.9
	Range	19.56	12.30	7.26	2.20	5.73	4.29	0.00	3.84	10.00
Residual Statistics	Max	3.94	2.96	3.94	0.69	1.60	0.81	0.07	0.90	0.92
	Min	-2.30	-2.30	-1.05	-0.80	-1.19	-0.96	0.07	-0.86	-1.79
	Mean[1]	0.04	0.11	-0.04	-0.06	0.15	-0.07	0.07	0.19	-0.13
	A-Mean[2]	0.59	1.08	0.48	0.37	0.64	0.23	0.07	0.39	0.61
	RMS[3]	0.85	1.30	0.77	0.45	0.75	0.33	0.07	0.50	0.79
	%RMS[4]	4.4%	10.6%	10.6%	20.4%	13.1%	7.7%		12.9%	7.9%



Summary Table - 90% Capture - V2

Observation Group	Name	ALL	Layer01	Layer02	Layer03	Layer04	Layer05	Layer07	Layer09	Layer13
	Count	215	53	58	8	25	34	1	27	9
Observation Summary	Max	-11.4	-11.4	-13.4	-17.8	-17.0	-17.9	-20.4	-18.9	-20.9
	Min	-30.9	-23.7	-20.7	-20.0	-22.8	-22.2	-20.4	-22.8	-30.9
	Range	19.56	12.30	7.26	2.20	5.73	4.29	0.00	3.84	10.00
Residual Statistics	Max	3.98	2.85	3.98	0.76	1.56	0.87	0.30	0.81	0.87
	Min	-2.26	-2.26	-1.12	-0.75	-1.17	-0.99	0.30	-0.67	-1.79
	Mean[1]	0.09	0.16	-0.02	0.05	0.18	0.06	0.30	0.21	-0.15
	A-Mean[2]	0.60	1.04	0.50	0.40	0.64	0.29	0.30	0.36	0.61
	RMS[3]	0.85	1.28	0.78	0.46	0.74	0.38	0.30	0.44	0.78
	%RMS[4]	4.3%	10.4%	10.8%	21.1%	12.9%	8.9%		11.5%	7.8%

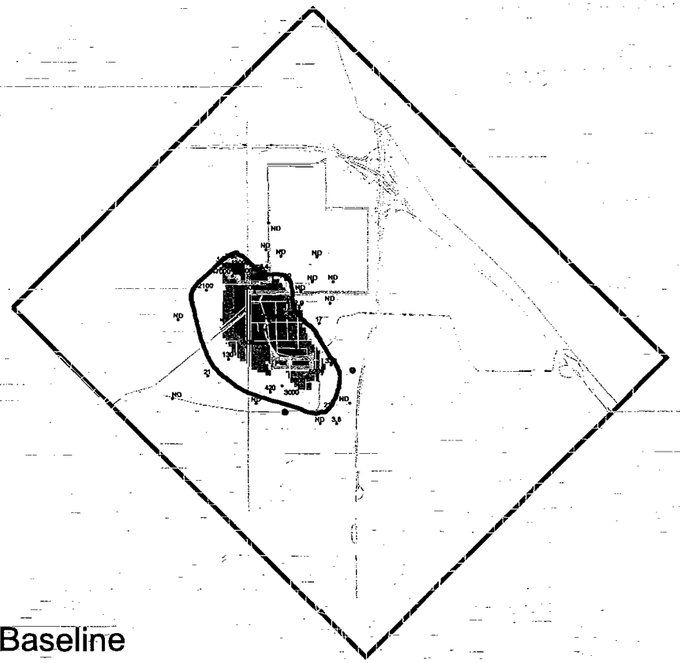


Summary Table - 75% Capture - V2

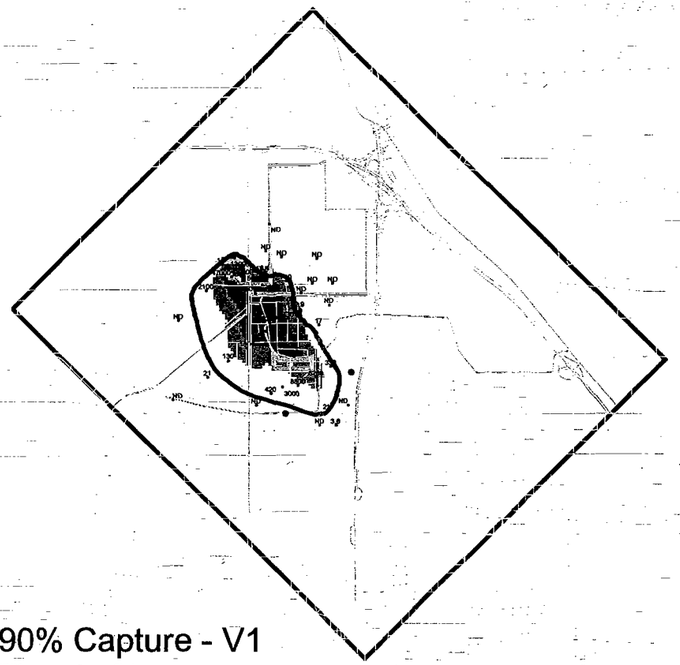
Observation Group	Name	ALL	Layer01	Layer02	Layer03	Layer04	Layer05	Layer07	Layer09	Layer13
	Count	215	53	58	8	25	34	1	27	9
Observation Summary	Max	-11.4	-11.4	-13.4	-17.8	-17.0	-17.9	-20.4	-18.9	-20.9
	Min	-30.9	-23.7	-20.7	-20.0	-22.8	-22.2	-20.4	-22.8	-30.9
	Range	19.56	12.30	7.26	2.20	5.73	4.29	0.00	3.84	10.00
Residual Statistics	Max	3.95	2.68	3.95	0.58	1.49	0.70	0.12	0.57	0.95
	Min	-1.82	-1.77	-1.01	-0.81	-1.29	-1.00	0.12	-0.68	-1.82
	Mean[1]	0.01	0.16	-0.06	-0.10	0.07	-0.15	0.12	0.06	-0.12
	A-Mean[2]	0.53	0.85	0.47	0.38	0.62	0.28	0.12	0.29	0.59
	RMS[3]	0.76	1.06	0.76	0.44	0.73	0.37	0.12	0.35	0.78
	%RMS[4]	3.9%	8.7%	10.5%	20.2%	12.8%	8.7%		9.2%	7.8%

Notes:
1. Mean Error
2. Absolute Mean Error
3. Root Mean Squared Error
4. Root Mean Squared Error normalized to observed water level range

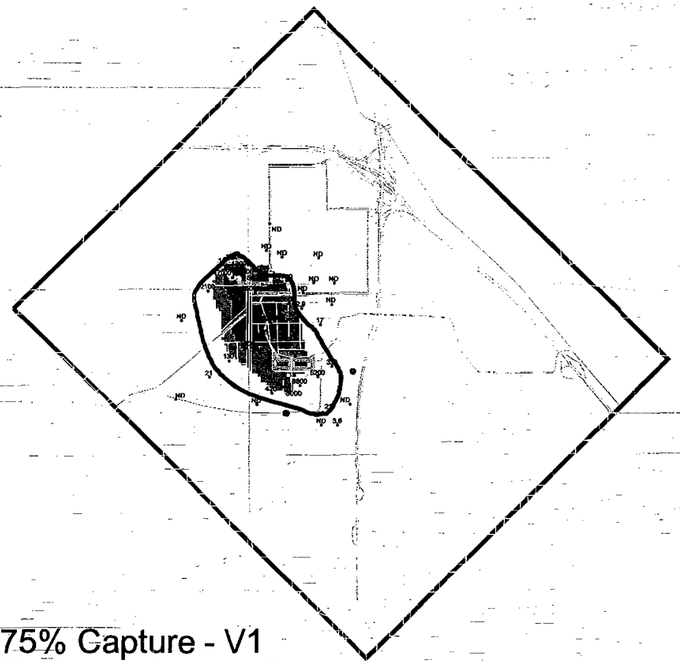
Figure 4-11
Comparison of Simulated vs. Measured Water Levels
Revised Baseline Calibration and Containment Failure Scenarios in the MBFC



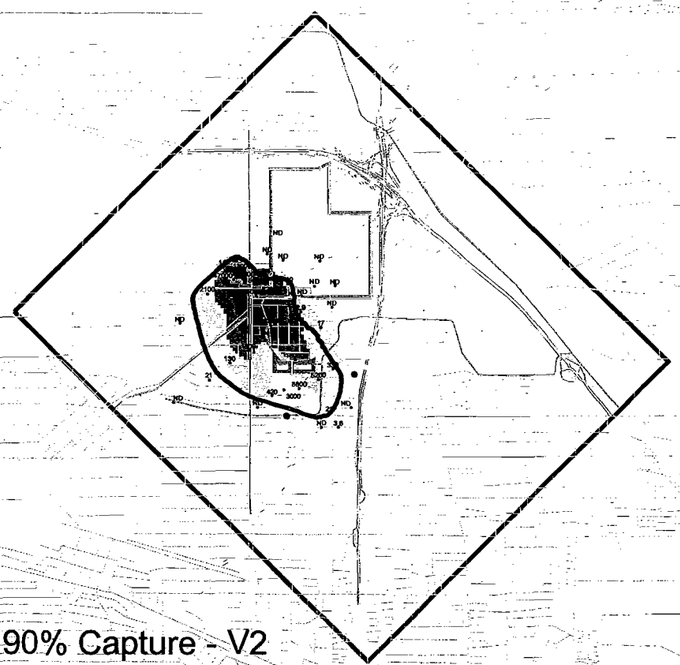
Baseline



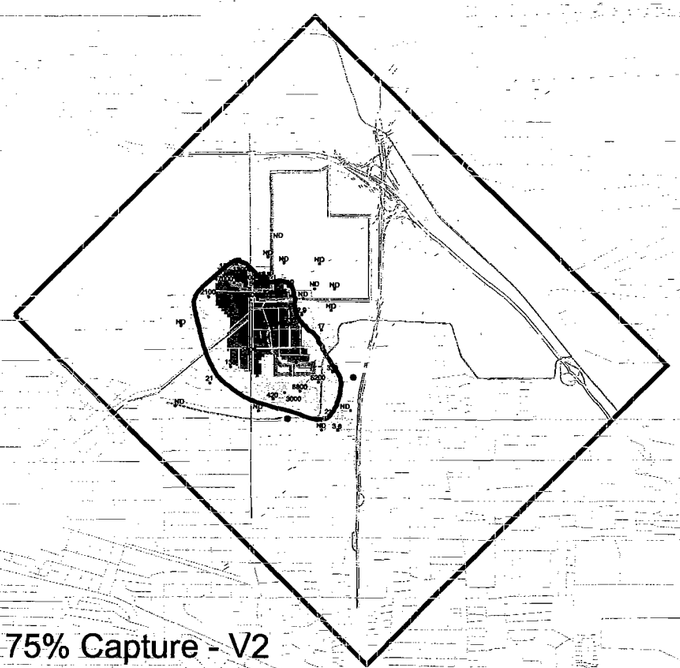
90% Capture - V1



75% Capture - V1



90% Capture - V2



75% Capture - V2

- Chlorobenzene Concentrations (µg/L), Janunary 2004
- Synthetic nondetect Chlorobenzene Concentration Points
- Interpreted 70 µg/L Contour, January 2004
- Streets
- Site Location
- Model Domain
- Base Map

Color Codes for Simulated Chlorobenzene Concentrations (µg/L)

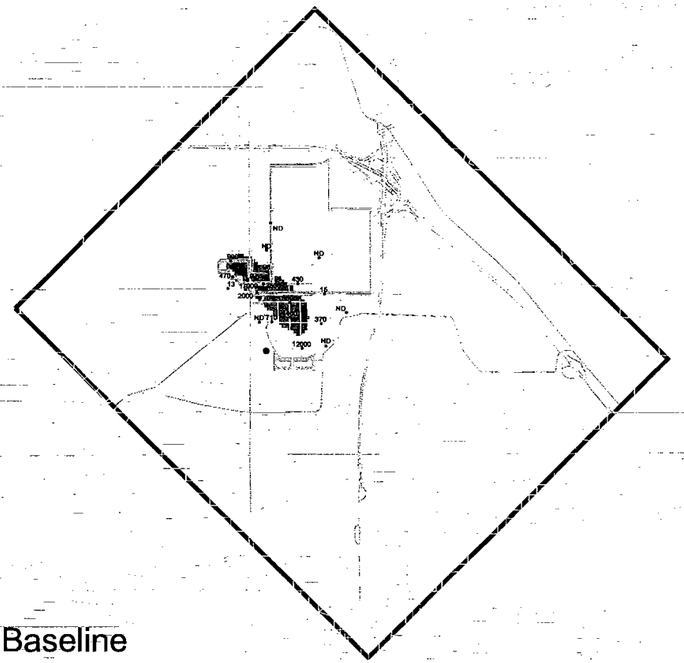
- 70 - 1000
- 1000 - 5000
- 5000 - 10000
- 10000 - 100000
- 100000 - 1000000

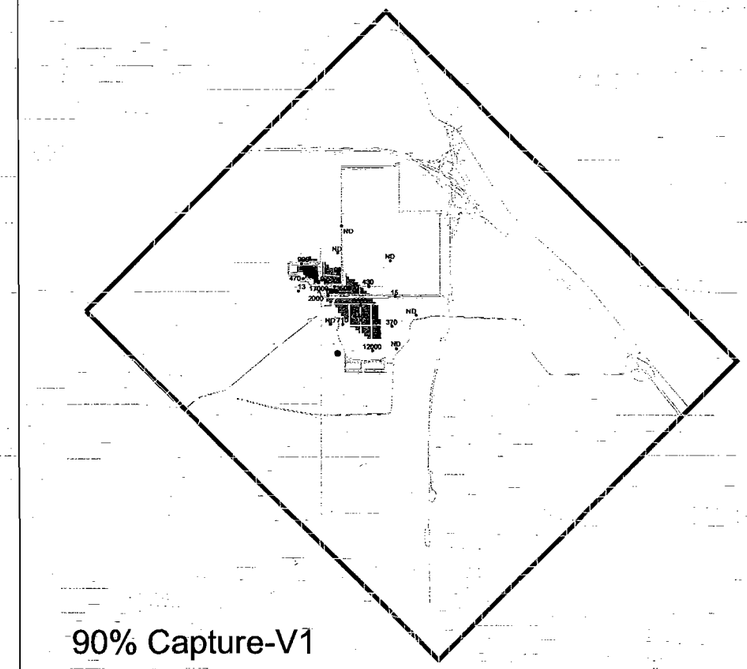
N

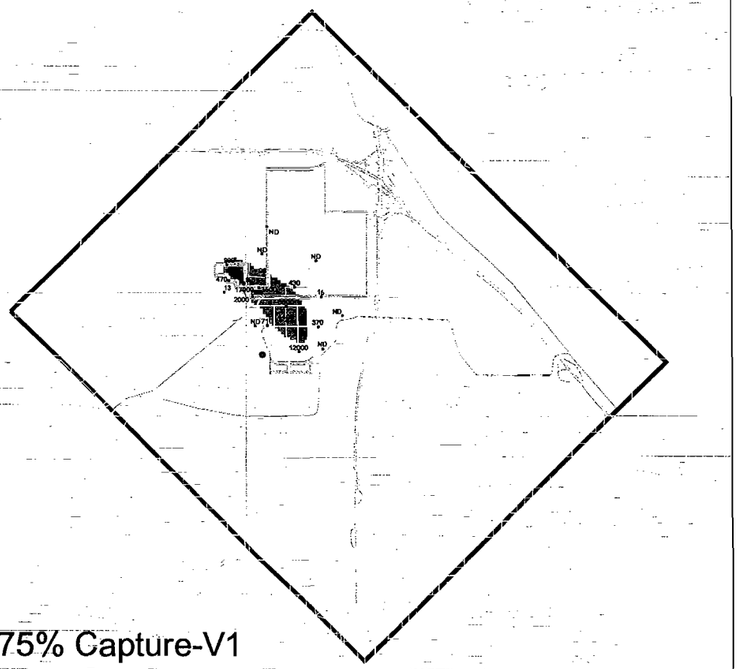
5000 Ft 5000 Ft

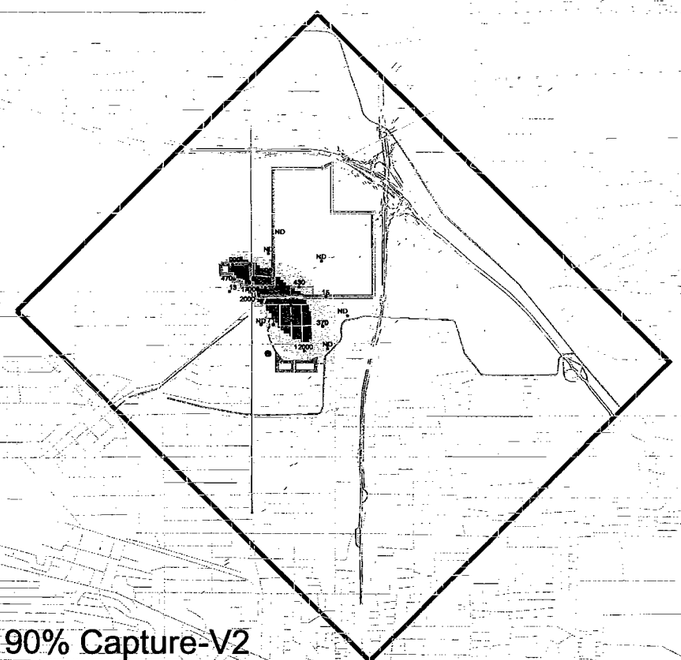
Figure 4-12
Comparison of Simulated Historical Chlorobenzene Plumes in the MBFC (1945-2004)
Revised Baseline Calibration and Containment Failure Scenarios in the MBFC

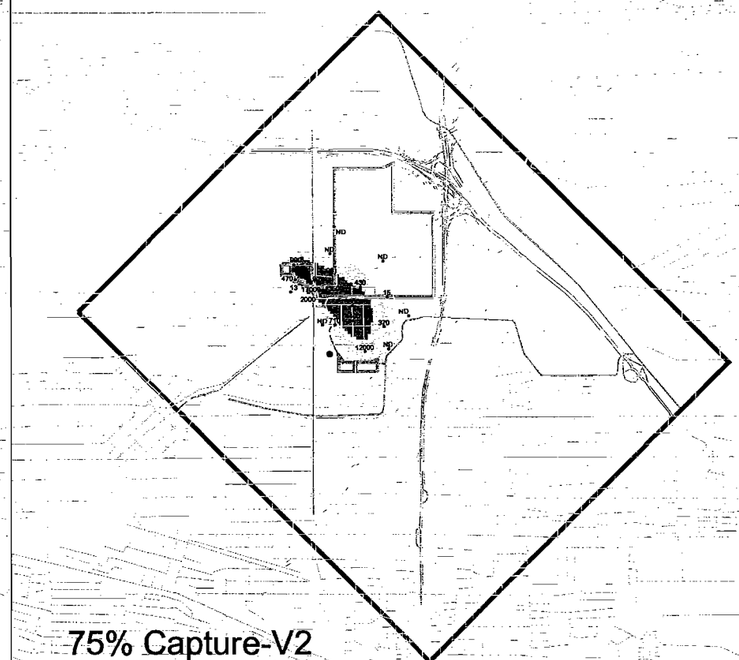
J:\Montrose_GW_Remedy\plots\Plume_Compare_Jun04.apr











- Chlorobenzene Concentrations (µg/L), Janunary 2004
- Synthetic nondetect Chlorobenzene Concentration Points

Streets
Site Location
Model Domain
Base Map

Color Codes for Simulated Chlorobenzene Concentrations (µg/L)

- 70 - 1000
- 1000 - 5000
- 5000 - 10000
- 10000 - 100000
- 100000 - 1000000

N

5000 Ft 5000 Ft

Figure 4-13
Comparison of Simulated Historical Chlorobenzene Plumes in the Gage (1945-2004)
Revised Baseline Calibration and Containment Failure Scenarios in the MBFC

J:\Montrose_GW_Remedy\plots\Plume_Compare_Jun04.apr

4.2 Assessment of Model Uncertainty for Predictions of Plume Containment

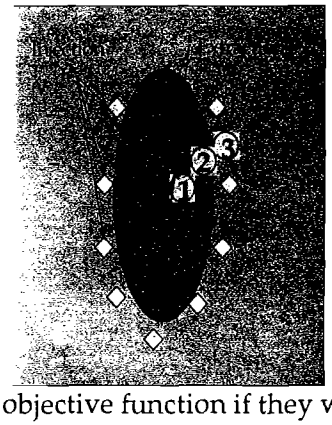
The ability of the remedial wellfield to contain the chlorobenzene plume is another important requirement of the ROD. The predictive uncertainty of the model with regard to containment was assessed to identify model parameters that are the most critical for reducing this uncertainty. This assessment was performed using the same methodology as that used for assessing predictive uncertainty with regard to mass and volume reduction described above. Several revisions to the initial baseline calibration were made prior to this analysis, based on input from reviewers. Containment targets, calibration results, model predictions with regard to plume containment, and the predictive uncertainty of the model are discussed below.

4.2.1 Containment Targets

Predictive targets addressing containment of the chlorobenzene plume in the target aquifers (MBFC and Gage Aquifer) were included in the calibration process. Predictive calibration of the model was performed to meet both calibration and predictive targets. As discussed above, the calibration targets included 1995 water levels, vertical water level differences, 1985 to 2004 chlorobenzene and p-CBSA data, and mass targets.

Predictive targets for the containment of the plume were developed using the MODPATH particle-tracking code. The modeling domain was subdivided into three zones including (1) within the chlorobenzene plume, (2) within 1,000 feet of the chlorobenzene plume, and (3) elsewhere within the modeling domain. A schematic diagram of these zones is shown in below. The chlorobenzene plume was defined based on a 70-µg/L contour from the Montrose 2004 Baseline Sampling report (Hargis, 2004). The 70-µg/L contour was "closed" south of well G-19 in the Gage Aquifer for the purposes of these analyses, although it was understood that the extent of the plume south of G-19 would have to be defined as part of the proposed data acquisition. Particles were placed in every cell of the model domain and labeled according to the zone that cell was located in. The predictive targets were set up in such a way that particles from Zones 1 and 2 (i.e., within, and in the immediate vicinity of, the chlorobenzene plume) resulted in the reduction of objective function if the particles were captured by extraction wells, and the increase of objective function if the particles reached other model boundaries. Particles from Zone 3 (i.e., elsewhere in the modeling domain) resulted in an increase of the objective function if they were captured by extraction wells.

Schematic Diagram of Containment Target Zones



A "counting weight" (i.e., contribution to objective function) was assigned to each particle based on a zone it was in, with weights being much higher for cells contained within the chlorobenzene plume than for cells in the other two zones. The total weighted count of

particles entering all extraction wells of the remedial wellfield was used as a means of assessing containment. Predictive targets of 90 percent and 75 percent containment were established for two different sets of runs to assess hydrogeologic conditions that could result in containment failure by the remedial wellfield, which would cause 10 percent and 25 percent, respectively, of the particles to escape containment.

As discussed below, sensitivities of predictive containment targets (i.e., the weighted count of particles) to each model parameter were calculated using PEST as part of the predictive calibration. As with the assessment of uncertainty for mass and volume reduction predictions, these sensitivities were used to estimate the contributions of various parameters to the predictive uncertainty associated with containment, and to identify model parameters that have the largest contribution (see Section 4.2.4 and Appendix A).

4.2.2 Calibration Results for Revised Baseline Calibration and Containment Failure Scenarios

4.2.2.1 Revisions to Baseline Calibration

As with the initial calibration, the revised baseline calibration is based solely on matching calibration targets and does not include predictive targets (i.e., the model was calibrated without regard to mass and volume reduction and/or capture of the plume). The following revisions were made to the original baseline calibration, based on the input provided by reviewers and to facilitate the predictive uncertainty analysis for plume containment:

- The error in source concentrations was corrected. Due to a data-entry error, PEST was assigning source concentrations that were lower than expected.
- Calibration targets for chlorobenzene concentrations above 500,000 µg/L measured in the source area wells were reduced to 500,000 µg/L to be consistent with the solubility of this constituent.
- Additional pilot points were added in the vicinity of the chlorobenzene plume to achieve better resolution for uncertainty analysis.
- Montrose aquifer test data provided by Hargis (April 5, 2004 fax) were added as fixed-value pilot points for four locations.
- The regularization target for hydraulic conductivity in the LBF was reduced from the original target used for the initial calibration, which was derived from the JGWFS model. As discussed in the Work Plan for Model Development (CH2M HILL, 2003), in the PEST calibration process, the calibrated values of hydraulic conductivity are estimated as close to the given regularization target as possible unless the deviation from this target is required to reproduce the observed conditions. The initial calibration indicated the lowest hydraulic conductivity in the LBF in the footprint of the plume, which is the area where the observed data are available for calibration. This suggested that lower hydraulic conductivity values in the LBF are required to match the measured and simulated results.
- 2004 baseline sampling data for chlorobenzene and p-CBSA were added to the calibration targets.

- Calibration weights at various wells were changed to improve the match in low-concentration wells at the western flank of the chlorobenzene plume in the Gage Aquifer.
- Three synthetic calibration targets with nondetect concentration values were added to the concentration calibration targets to preclude PEST calibration from simulating chlorobenzene migration in the areas where it is not believed to be present based on the current interpretation of existing data (see Figures 4-12 and 4-13). One target was assigned to the Gage Aquifer west of the chlorobenzene plume (x-coord = 4197790, y-coord = 4054550). This nondetect concentration target was added to the calibration process to preclude PEST from simulating leakage of elevated chlorobenzene concentrations from the high-concentration plume in the MBFC in the vicinity of well BF-21 into the area west of the currently defined Gage Aquifer plume. It is important to note, however, that although a nondetect target was added to this area for the purposes of the data gap analysis, the potential leakage from the MBFC west of the currently defined plume in the Gage Aquifer needs to be assessed, because this area is located downgradient of nondetect well G-16, which currently delineates the western edge of the chlorobenzene plume in the Gage Aquifer. Two other nondetect concentration targets were assigned to the MBFC between wells BF-30 and BF-27 (x-coord = 4198473, y-coord = 4052473), and wells BF-12 and BF-28 (x-coord = 4200729, y-coord = 4053875). These targets were added to the calibration process to preclude PEST from overestimating the extent of the chlorobenzene plume in the MBFC, between and downgradient of monitoring wells BF-30, BF-27, and BF-28. While no calibration targets (i.e., wells) exist in this area to contribute to objective function and preclude PEST from simulating narrow high-concentration distributions escaping between low-concentration wells, such distributions do not appear to be realistic. PEST overestimates the downgradient extent of the chlorobenzene plume in these locations and attempts to "squeeze" the plume between the low-concentration wells because, as in the initial baseline calibration, it is unable to reproduce the relatively steep concentration gradients between high-concentration wells BF-16, BF- 11, and BF-17 and low-concentration wells BF-30, BF-27, BF-25, and BF-28 at the toe of the chlorobenzene plume. As discussed above, these steep concentration gradients could be attributed to variable strength and timing of contaminant sources, local and regional variation in the historical flow field and hydrogeologic and physical properties of the aquifer formation, and other unknown factors that could not be considered in the calibration process.

4.2.2.2 Failure Scenarios with Regard to Plume Containment

Four scenarios were generated to assess combinations of model parameters that could result in failure of chlorobenzene plume containment. Failure scenarios with regard to the plume containment could not be developed for the Gage Aquifer, however. This indicates that given the existing data and assumed extent of the Gage plume (i.e., the assumed 70-µg/L contour for the chlorobenzene plume in the Gage Aquifer), the certainty of containing this plume with the JGWFS wellfield is fairly high. The simulated failure scenarios for the plume containment in the MBFC are described below.

1. **90 percent failure scenario with respect to plume containment in the MBFC – Version 1:** This failure scenario was generated using the same calibration targets as the revised baseline calibration, and also included a predictive target for not

capturing 10 percent of the chlorobenzene plume in the MBFC. The distribution of weights for calibration targets for this solution was the same as that for the revised baseline calibration.

2. **75 percent failure scenario with respect to plume containment in the MBFC – Version 1:** This failure scenario was generated using the same calibration targets as the revised baseline calibration, and also included a predictive target for not capturing 25 percent of the chlorobenzene plume in the MBFC. The distribution of weights for calibration targets for this solution was the same as that for baseline calibration.

3. **90 percent failure scenario with respect to plume containment in the MBFC – Version 2:** This failure scenario was generated using the same calibration targets as the baseline calibration, and also included a predictive target for not capturing 10 percent of the chlorobenzene plume in the MBFC. The distribution of weights for this scenario was altered compared to Version 1 by increasing weights of low-concentration and nondetect wells delineating the extent of the plume in the MBFC.

4. **75 percent failure scenario with respect to plume containment in the MBFC – Version 2:** This failure scenario was generated using the same calibration targets as the baseline calibration, and also included a predictive target for not capturing 25 percent of the chlorobenzene plume in the MBFC. The distribution of weights for this scenario was altered compared to Version 1 by increasing weights of low-concentration and nondetect wells delineating the extent of the plume in the MBFC.

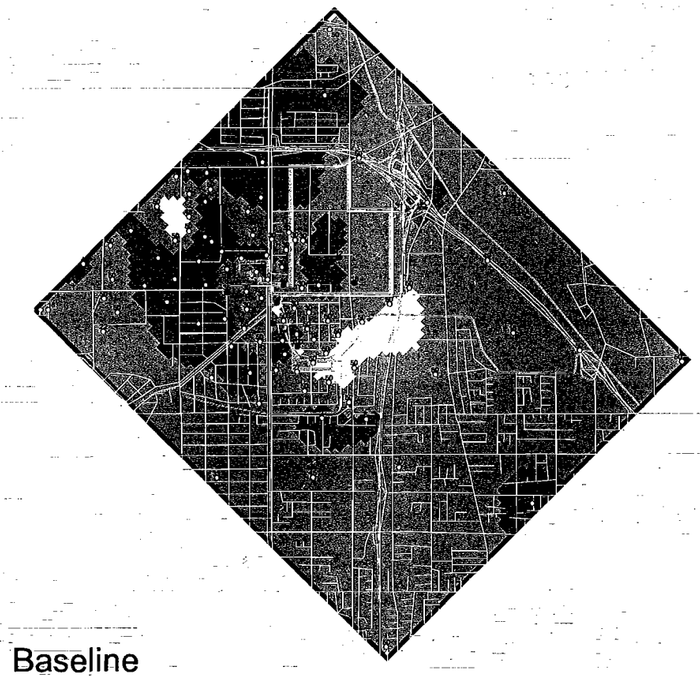
4.2.2.3 Calibration Results

Figure 4-11 provides a comparison of head calibration statistics for the revised calibration and four failure scenarios in the MBFC. Figures 4-12 and 4-13 provide a comparison of calibrated historical chlorobenzene plumes for these scenarios. As indicated on the figures, the flow calibration for the five models is comparable with slightly better fit for the baseline calibration and 75 percent failure scenario – Version 2. Transport calibration, while comparable for the Gage Aquifer, appears to be better for the baseline calibration in the MBFC as opposed to the failure scenarios. Version 1 failure scenarios overestimate downgradient migration of the low-concentration chlorobenzene plume in the MBFC, while generally matching high concentrations. Version 2 failure scenarios underestimate migration of the high-concentration plume in the MBFC, while generally matching low concentrations. These differences in transport calibration between failure scenarios for Versions 1 and 2 are consistent with the differences in calibration weights between the two scenarios (i.e., Version 2 scenarios have higher weights for low concentration wells).

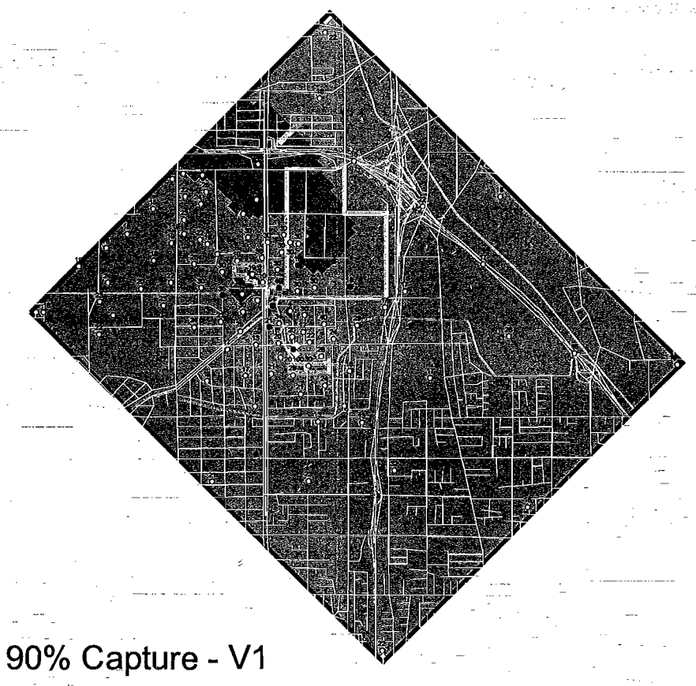
Figures 4-14 through 4-19 show the comparison of the calibrated horizontal hydraulic conductivity distribution for the revised baseline calibration and four failure scenarios in the MBFB, MBFC, LBF, Gage Aquifer, GLA, and Lynwood Aquifer. Figure 4-20 shows the comparison of the calibrated recharge for these scenarios and Figure 4-21 shows the comparison of model transport parameters.

Based on the comparison of the calibrated distributions of model parameters, the greatest differences were observed in the distribution of horizontal hydraulic conductivity in the MBFC. A high-conductivity zone (i.e., 730 feet per day [ft/day]) was created by PEST at the

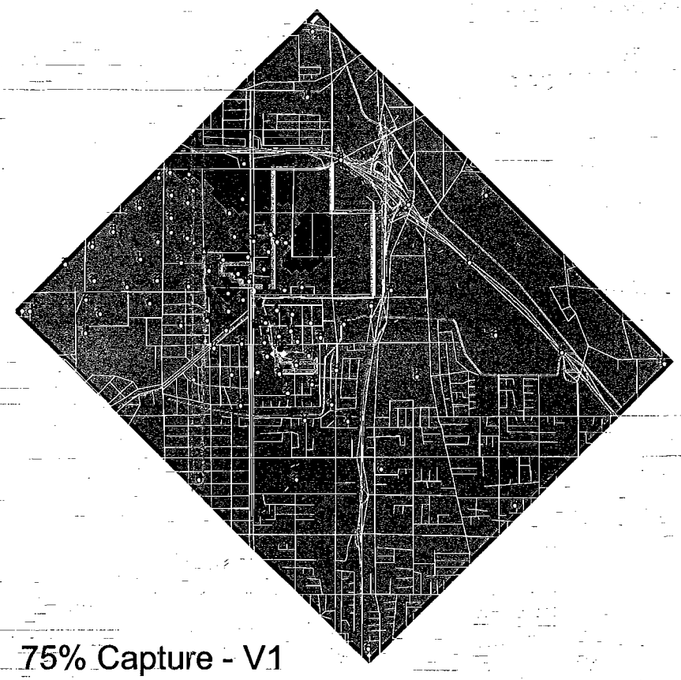
toe of the chlorobenzene plume for all failure scenarios, while hydraulic conductivities for the baseline calibration ranged at that location from about 30 to 270 ft/day. Transport parameters were generally similar for all scenarios, except for the values of porosity, which were higher in the MBFC for all failure scenarios (i.e., 0.4), as opposed to porosity of 0.3 for this unit in the baseline calibration. The higher values of porosity generally result is slower velocity of contaminant migration. It is likely that PEST assigned higher porosity to the MBFC because it was required to compensate for the high hydraulic conductivity at the toe of the plume, which otherwise would cause an overestimate of the extent of the historical plume in the MBFC.



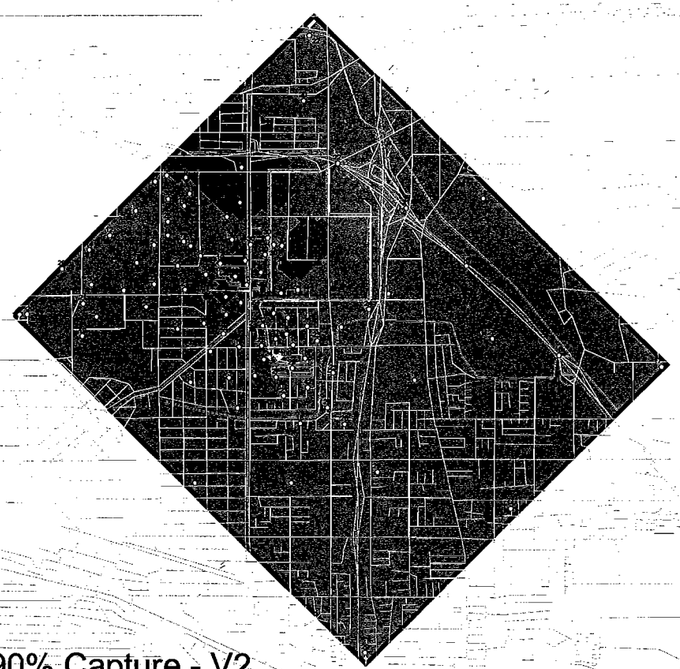
Baseline



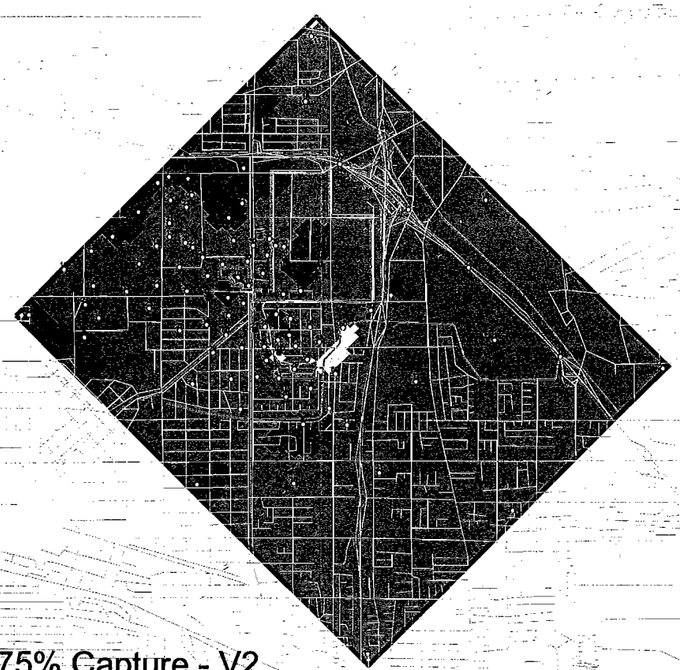
90% Capture - V1



75% Capture - V1



90% Capture - V2



75% Capture - V2

- Pilot Points, Values from Model Calibration
- Pilot Points, Values from Aquifer Tests

Streets

Site Location

Model Domain

Base Map

Color Codes for Horizontal Hydraulic Conductivity (ft/day)

- 0 - 0.001
- 0.001 - 0.01
- 0.01 - 0.1
- 0.1 - 1
- 1 - 10
- 10 - 50
- 50 - 100
- 100 - 300
- 300 - 500
- 500 - 1000

N

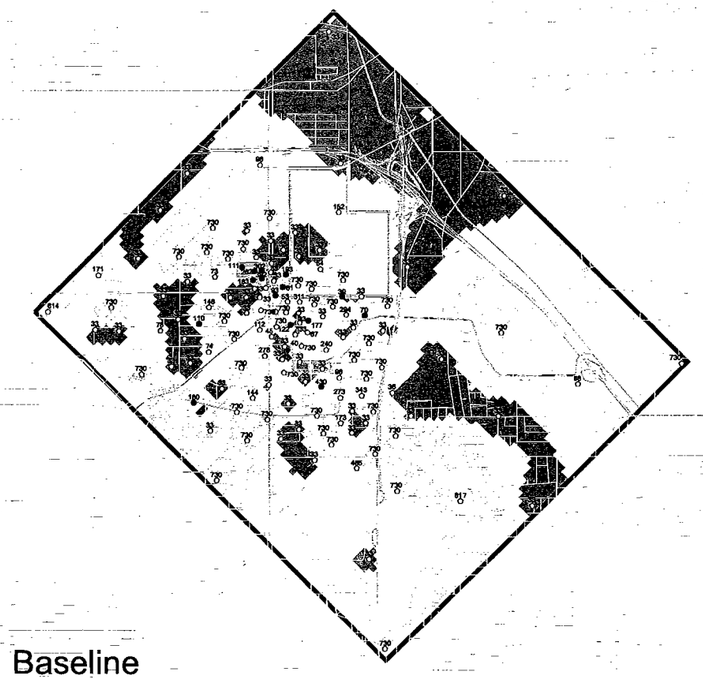
5000 Ft 5000 Ft

Figure 4-14

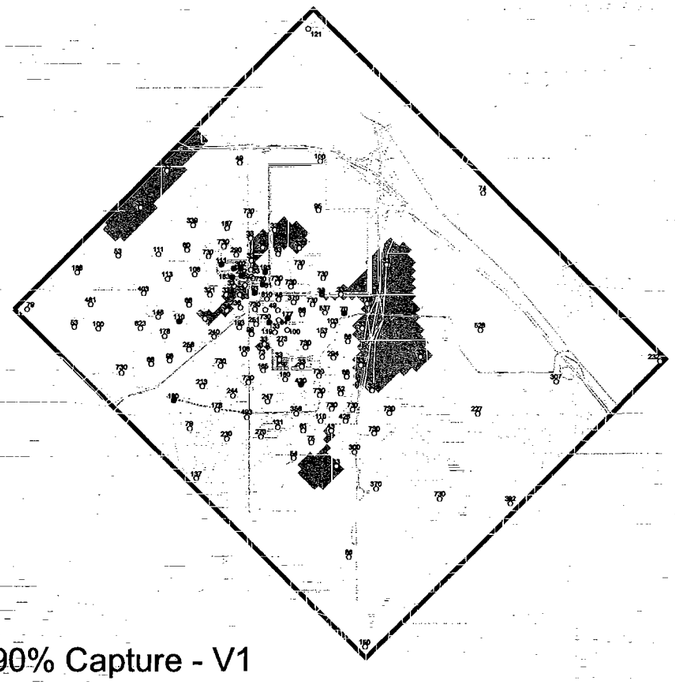
Comparison of Calibrated Horizontal Hydraulic Conductivity in the MBFB

Revised Baseline Calibration and Containment Failure Scenarios in the MBFC

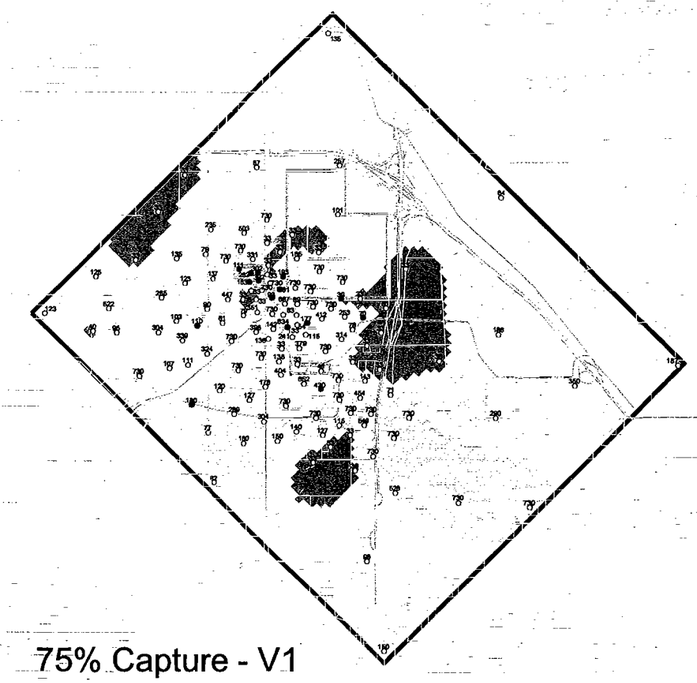
J:\Montrose_GW_Remedy\Plots\Kx_Compare_Jun04.apr



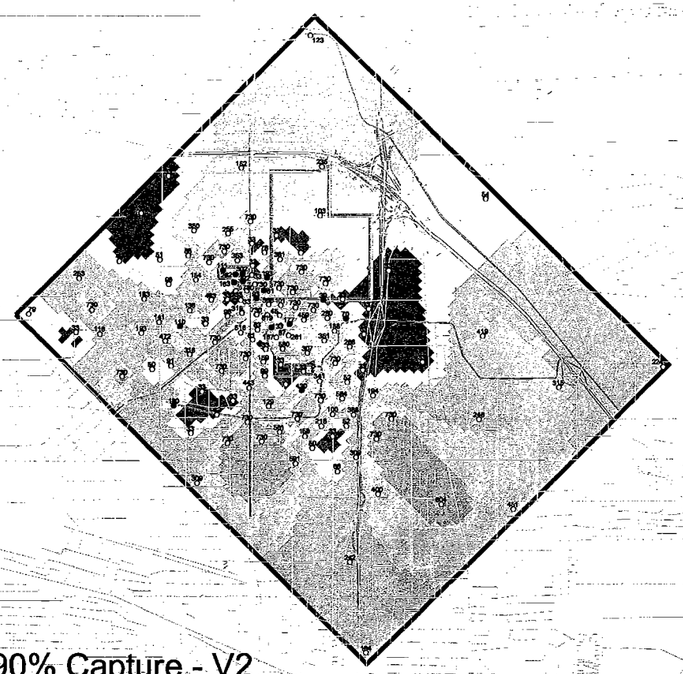
Baseline



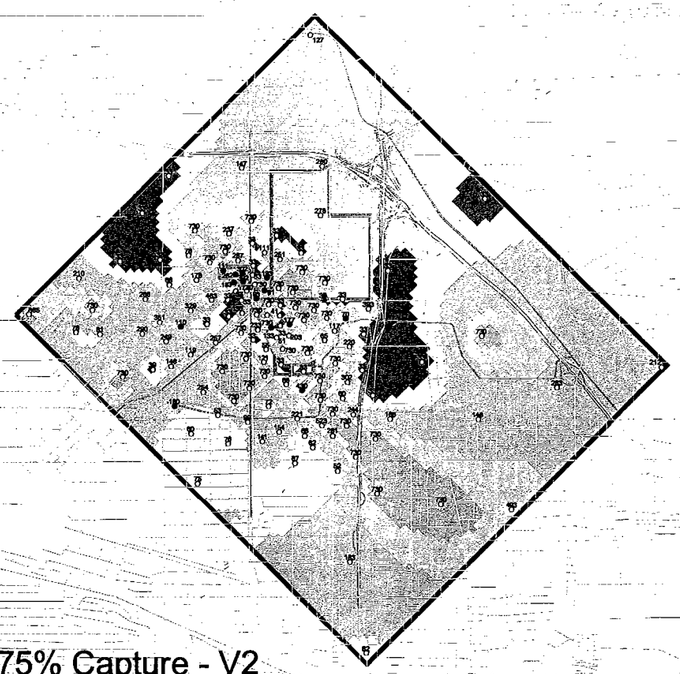
90% Capture - V1



75% Capture - V1



90% Capture - V2



75% Capture - V2

- Pilot Points, Values from Model Calibration
- Pilot Points, Values from Aquifer Tests
- Streets
- Site Location
- Model Domain
- Base Map

Color Codes for Horizontal Hydraulic Conductivity (ft/day)

- 0 - 0.001
- 0.001 - 0.01
- 0.01 - 0.1
- 0.1 - 1
- 1 - 10
- 10 - 50
- 50 - 100
- 100 - 300
- 300 - 500
- 500 - 1000

N

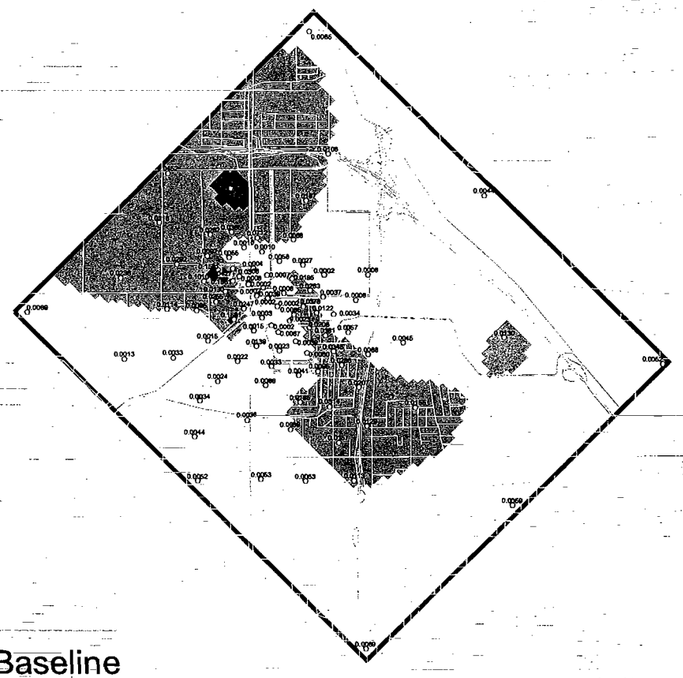
5000 Ft 5000 Ft

Figure 4-15

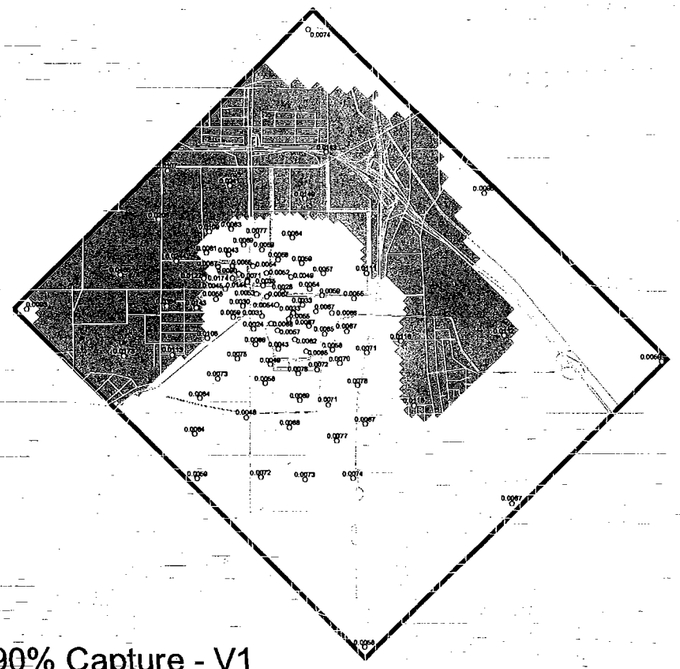
Comparison of Calibrated Horizontal Hydraulic Conductivity in the MBFC

Revised Baseline Calibration and Containment Failure Scenarios in the MBFC

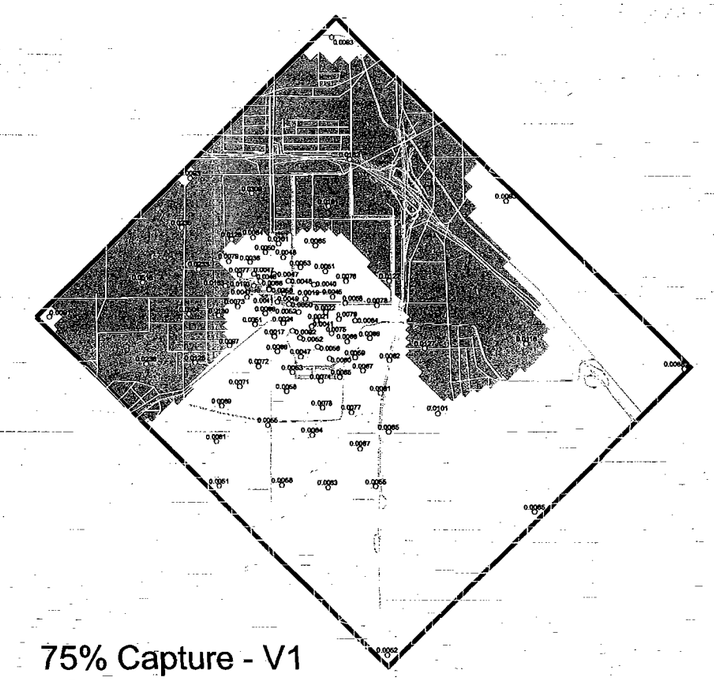
J:\Montrose_GW_Remedy\Plots\Kx_Compare_Jun04.apr



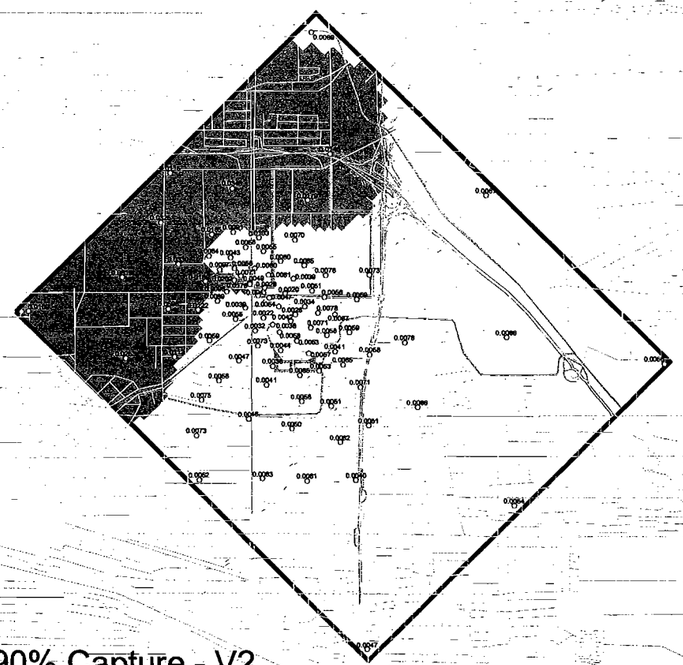
Baseline



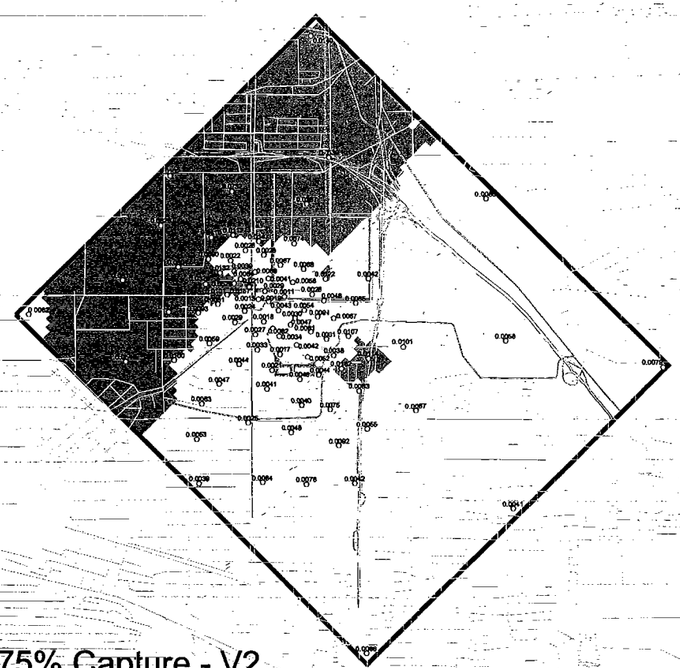
90% Capture - V1



75% Capture - V1



90% Capture - V2



75% Capture - V2

- Pilot Points, Values from Model Calibration
- Pilot Points, Values from Aquifer Tests

Streets

Site Location

Model Domain

Base Map

Color Codes for Horizontal Hydraulic Conductivity (ft/day)

- 0 - 0.001
- 0.001 - 0.01
- 0.01 - 0.1
- 0.1 - 1
- 1 - 10
- 10 - 50
- 50 - 100
- 100 - 300
- 300 - 500
- 500 - 1000

N

5000 Ft 5000 Ft

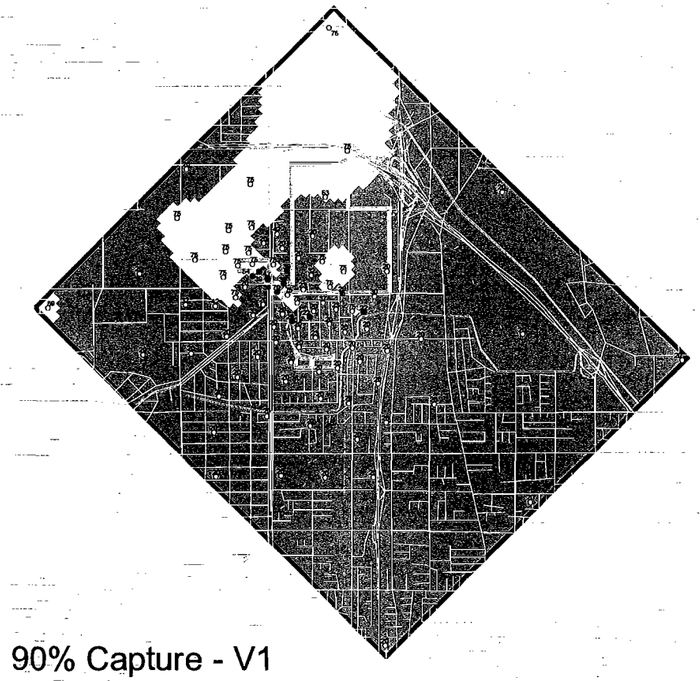
Figure 4-16

Comparison of Calibrated Horizontal Hydraulic Conductivity in the LBF

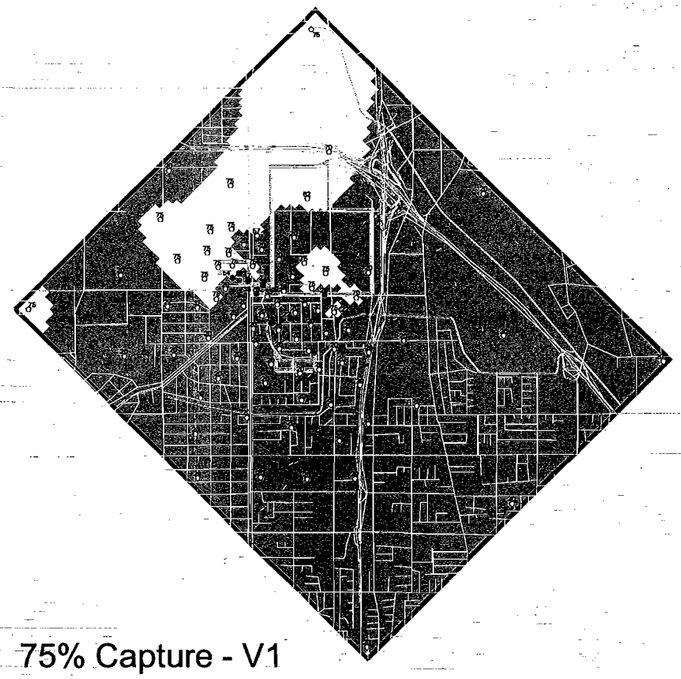
Revised Baseline Calibration and Containment Failure Scenarios in the MBFC

J:\Montrose_GW_Remedy\Plots\Kx_Compare_Jun04.apr

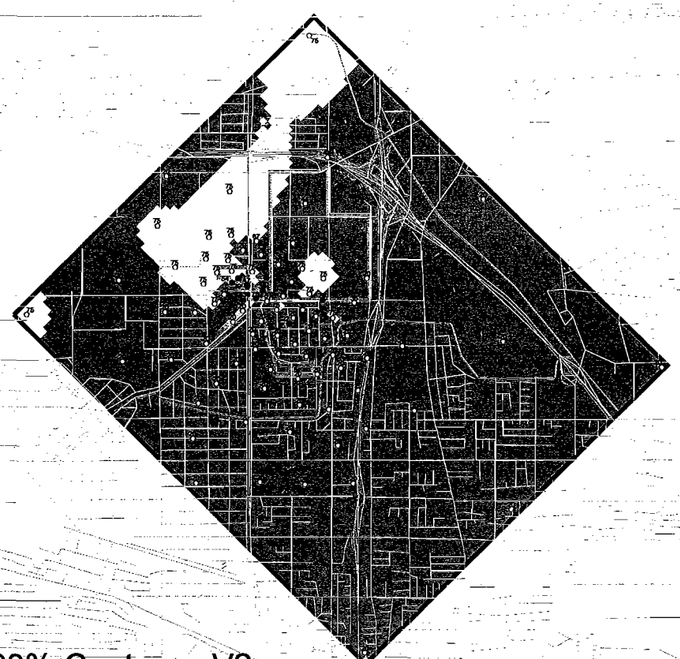
Baseline



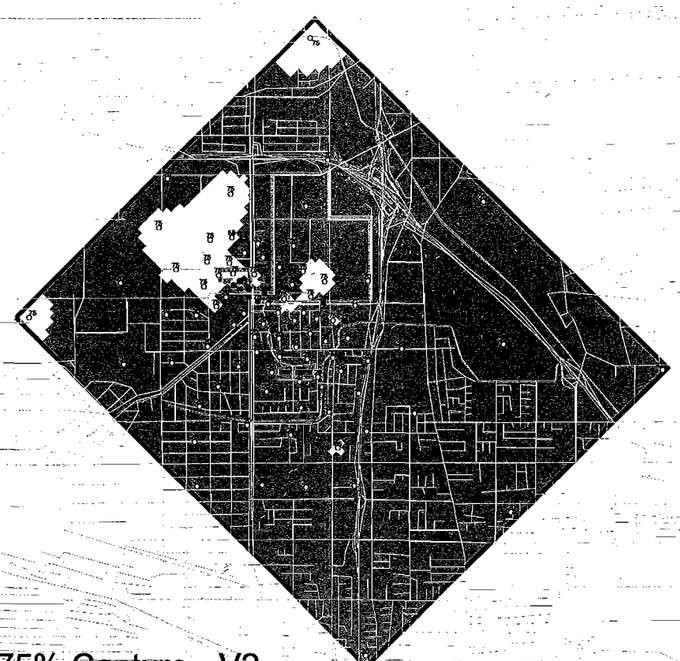
90% Capture - V1



75% Capture - V1



90% Capture - V2



75% Capture - V2

- Pilot Points, Values from Model Calibration
- Pilot Points, Values from Aquifer Tests
- Streets
- Site Location
- Model Domain
- Base Map

Color Codes for Horizontal Hydraulic Conductivity (ft/day)

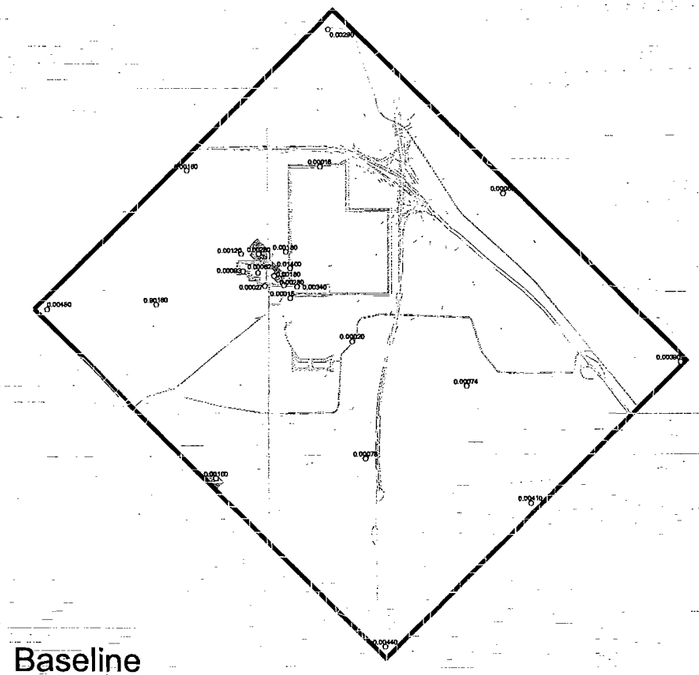
- 0 - 0.001
- 0.001 - 0.01
- 0.01 - 0.1
- 0.1 - 1
- 1 - 10
- 10 - 50
- 50 - 100
- 100 - 300
- 300 - 500
- 500 - 1000

N

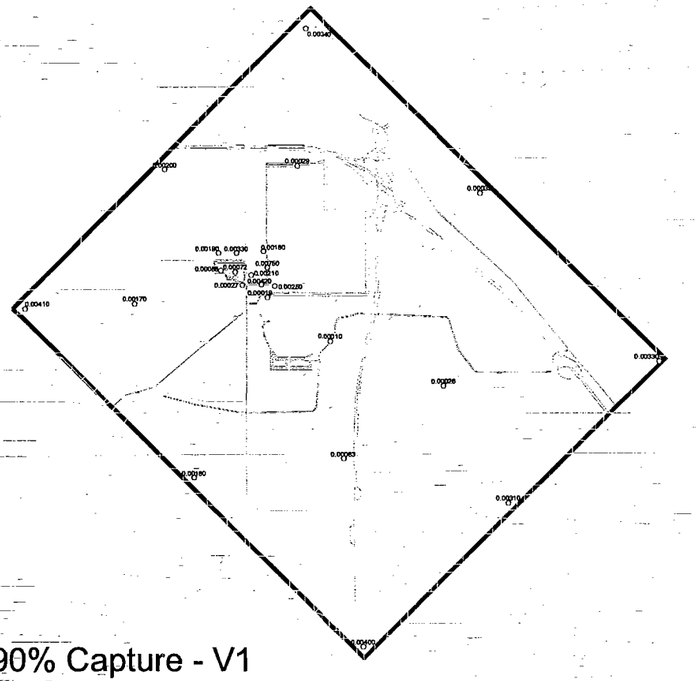
5000 Ft 5000 Ft

Figure 4-17
Comparison of Calibrated Horizontal Hydraulic Conductivity in the Gage
Revised Baseline Calibration and Containment Failure Scenarios in the MBFC

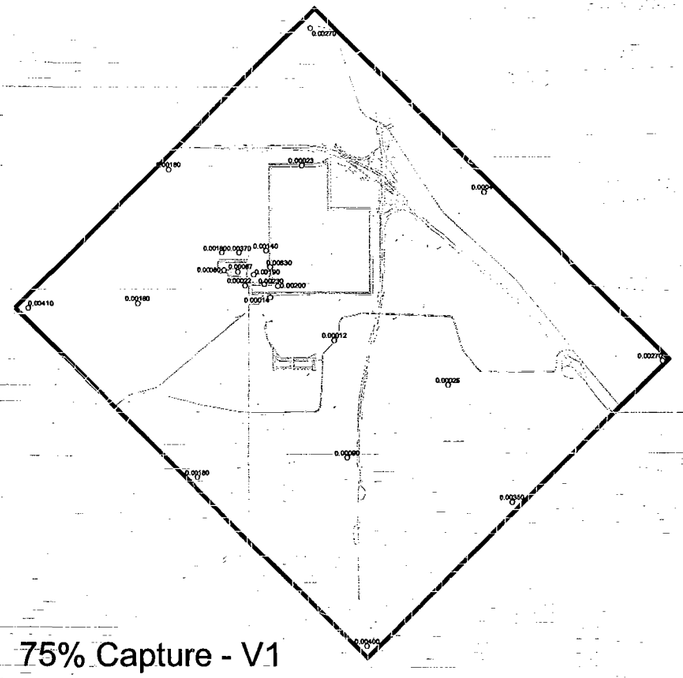
J:\Montrose_GW_Remedy\Plots\Kx_Compare_Jun04.apr



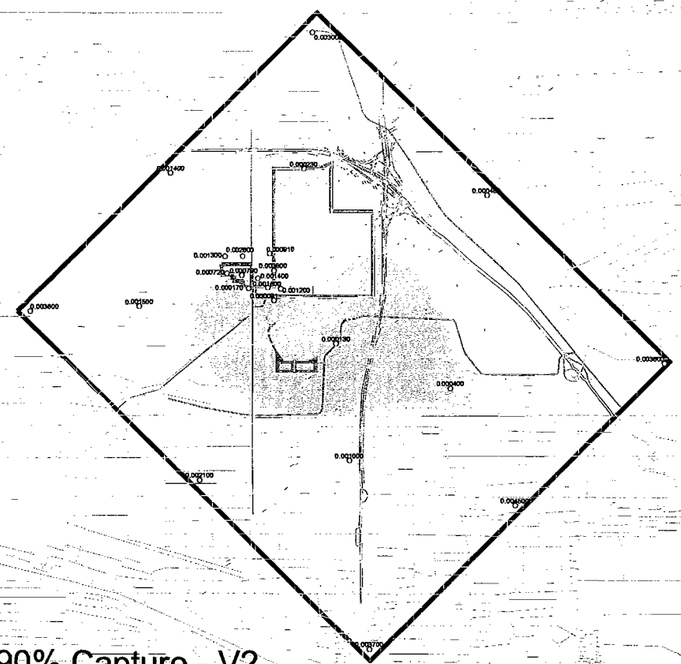
Baseline



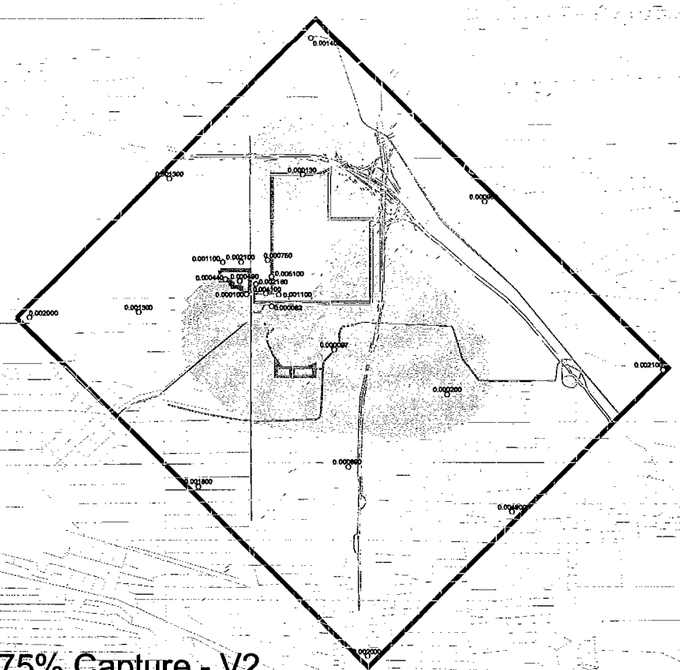
90% Capture - V1



75% Capture - V1



90% Capture - V2



75% Capture - V2

- Pilot Points, Values from Model Calibration
- Pilot Points, Values from Aquifer Tests

Streets

Site Location

Model Domain

Base Map

Color Codes for Horizontal Hydraulic Conductivity (ft/day)

- 0 - 0.001
- 0.001 - 0.01
- 0.01 - 0.1
- 0.1 - 1
- 1 - 10
- 10 - 50
- 50 - 100
- 100 - 300
- 300 - 500
- 500 - 1000

N

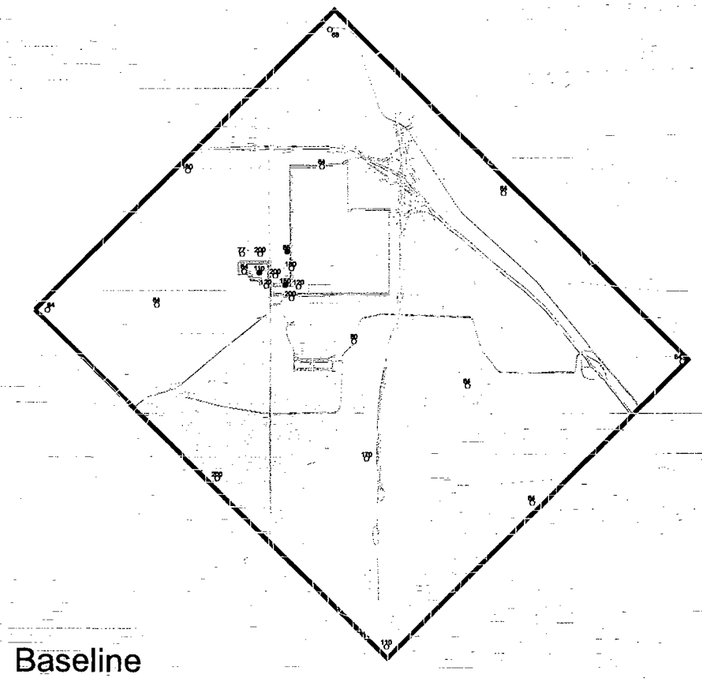
5000 Ft 5000 Ft

Figure 4-18

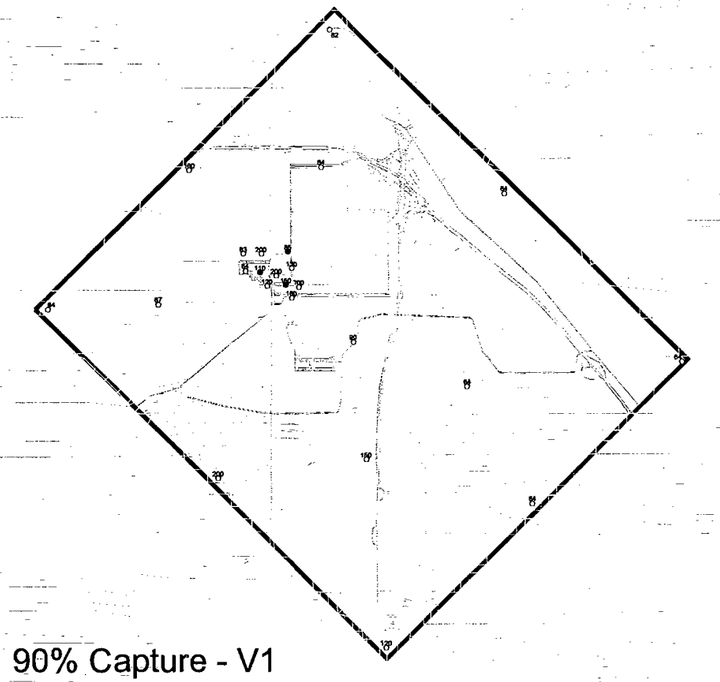
Comparison of Calibrated Horizontal Hydraulic Conductivity in the GLA

Revised Baseline Calibration and Containment Failure Scenarios in the MBFC

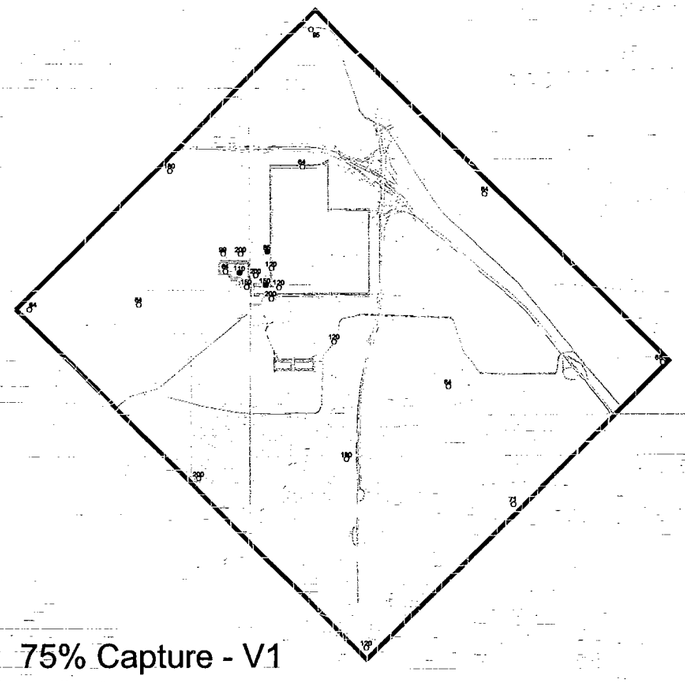
J:\Montrose_GW_Remedy\Plots\Kx_Compare_Jun04.apr



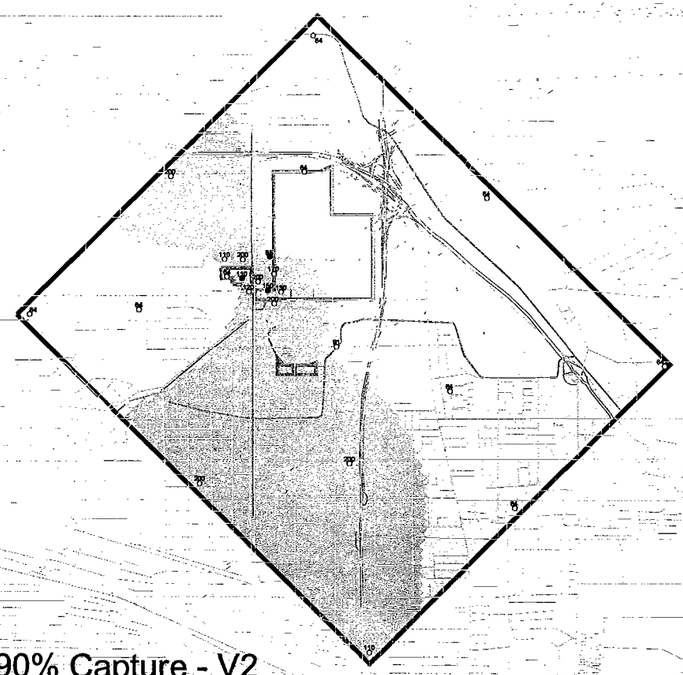
Baseline



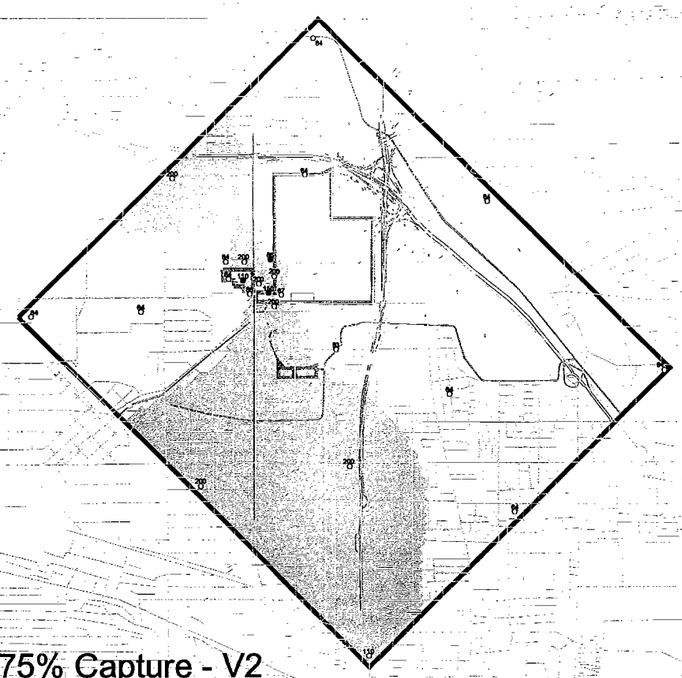
90% Capture - V1



75% Capture - V1



90% Capture - V2



75% Capture - V2

- Pilot Points, Values from Model Calibration
- Pilot Points, Values from Aquifer Tests
- Streets
- Site Location
- Model Domain
- Base Map

Color Codes for Horizontal Hydraulic Conductivity (ft/day)

- 0 - 0.001
- 0.001 - 0.01
- 0.01 - 0.1
- 0.1 - 1
- 1 - 10
- 10 - 50
- 50 - 100
- 100 - 300
- 300 - 500
- 500 - 1000

N

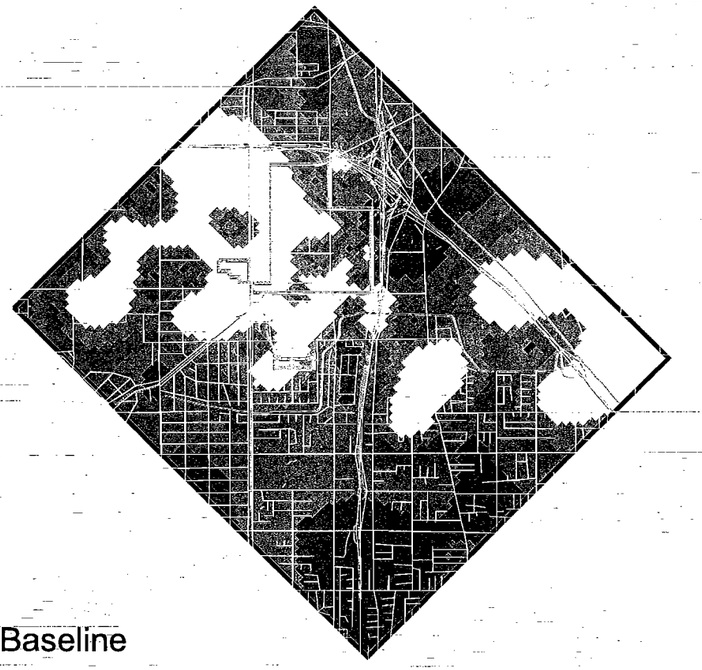
5000 Ft 5000 Ft

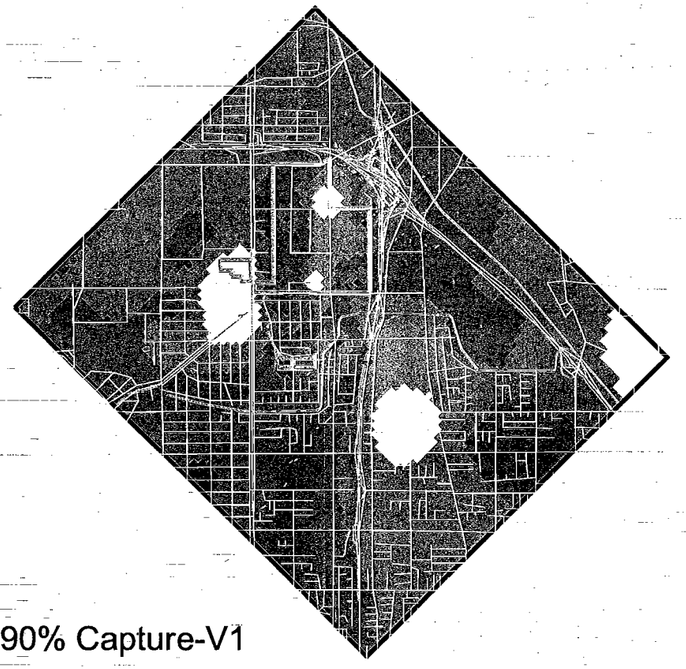
Figure 4-19

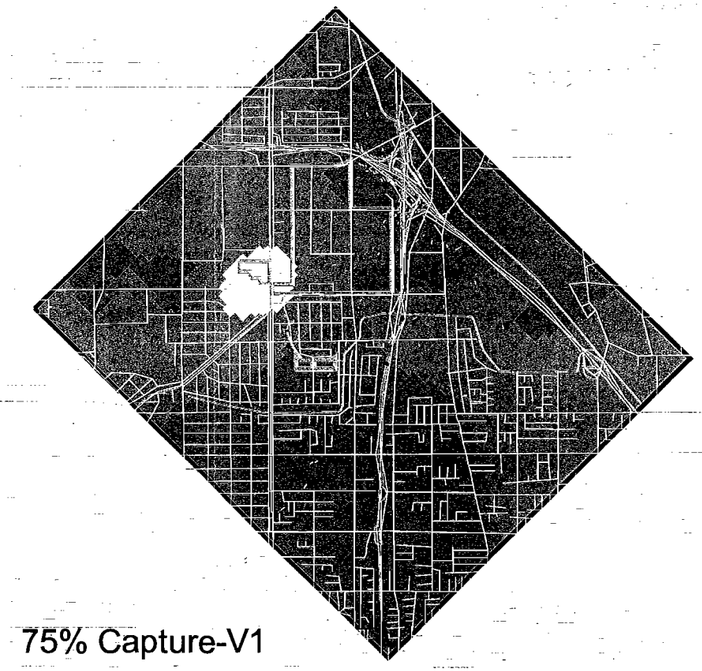
Comparison of Calibrated Horizontal Hydraulic Conductivity in the Lynwood

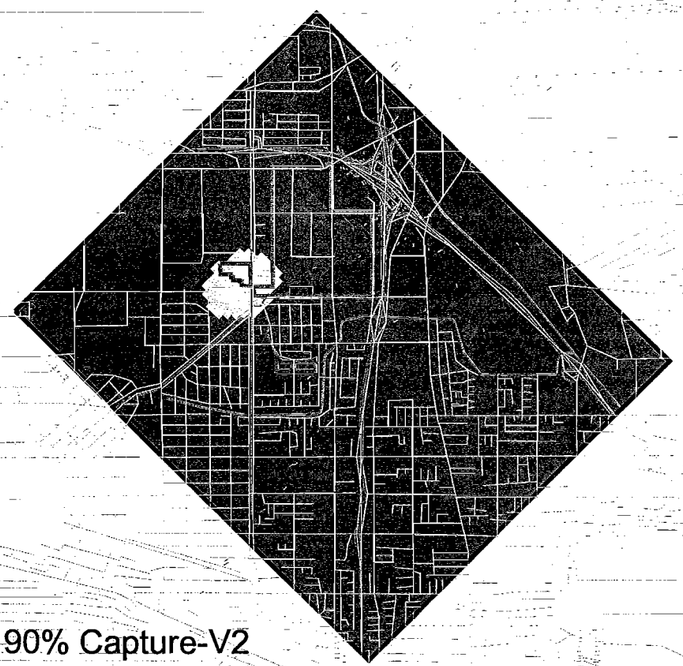
Revised Baseline Calibration and Containment Failure Scenarios in the MBFC

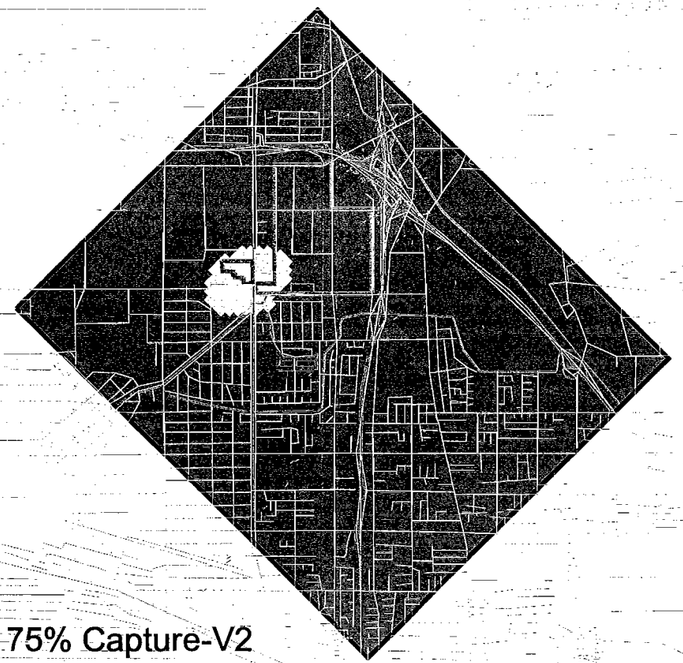
J:\Montrose_GW_Remedy\Plots\Kx_Compare_Jun04.apr











Streets
Site Location
Model Domain
Base Map

Color Codes for Surface Recharge (Inches/Year)

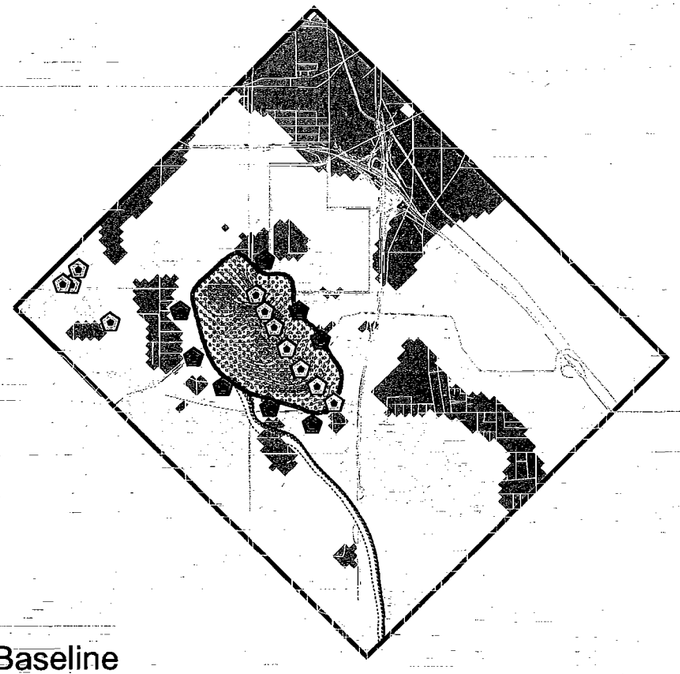
- 0 - 0.2
- 0.2 - 0.4
- 0.4 - 0.6
- 0.6 - 0.8
- 0.8 - 1
- 1 - 1.2
- 1.2 - 1.4
- 1.4 - 1.6
- 1.6 - 1.8
- 1.8 - 2

N

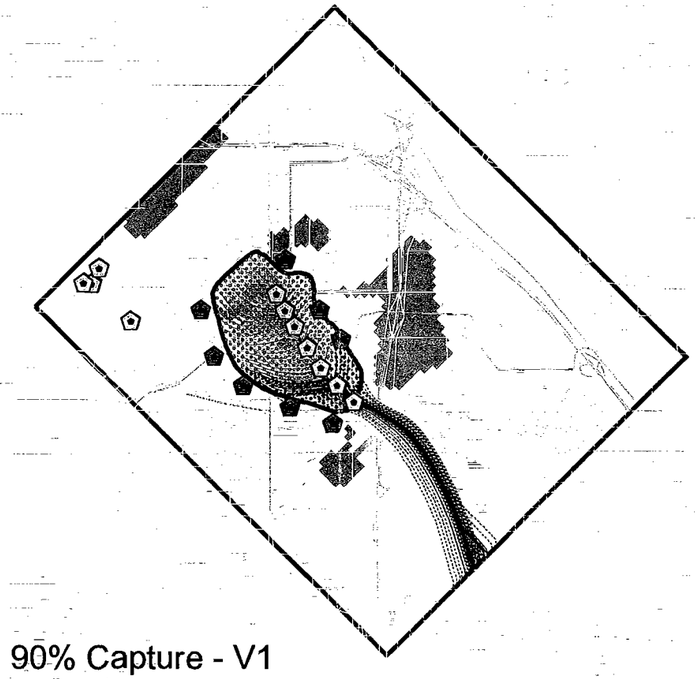
5000 Feet 5000 Feet

Figure 4-20
Comparison of Calibrated Surface Recharge
Revised Baseline Calibration and
Containment Failure Scenarios in the MBFC

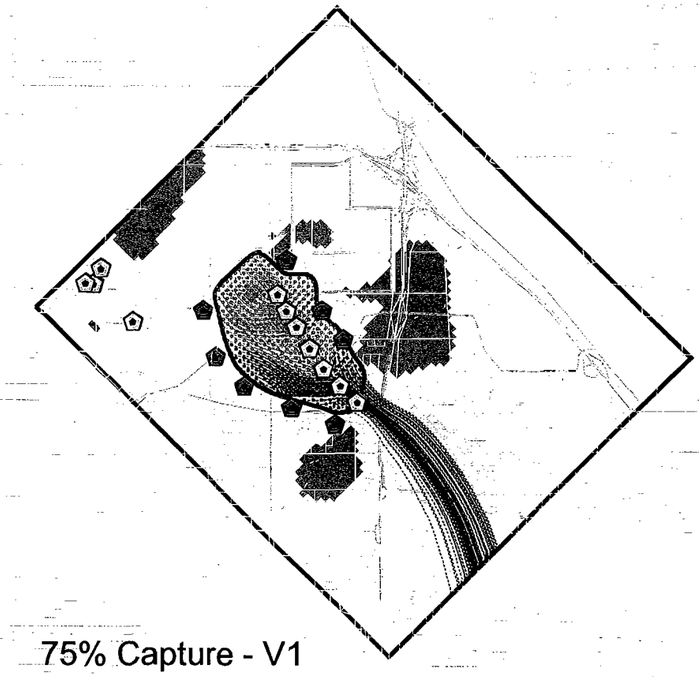
J:\Montrose_GW_Remedy\plots\Recharge_Compare_Jun04.apr



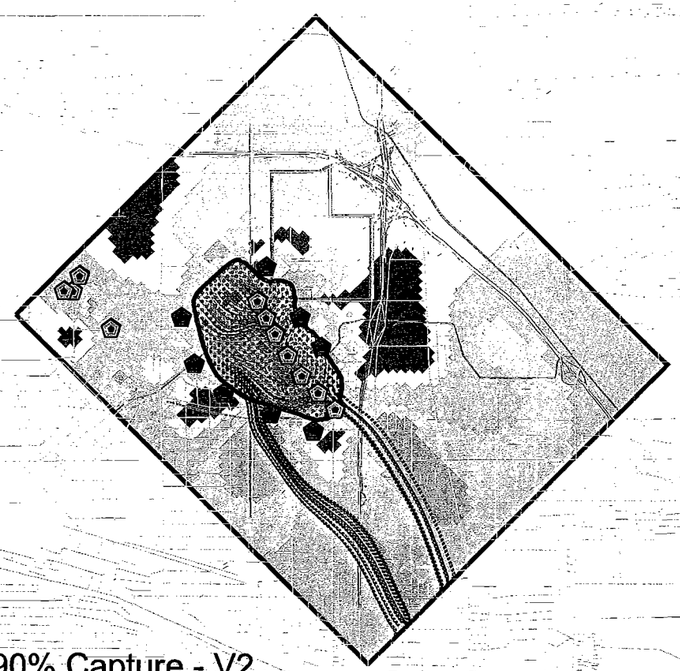
Baseline



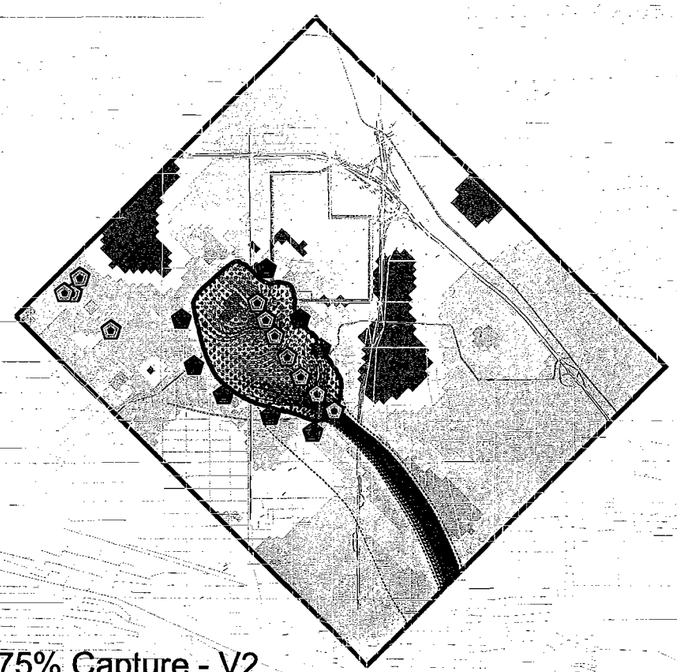
90% Capture - V1



75% Capture - V1



90% Capture - V2



75% Capture - V2

Particle Starting Locations, Captured by Extraction Wells
Particle Starting Locations, Not Captured by Extraction Wells
Interpreted 70 µg/L Concentration Contour, January 2004
Particle Pathlines, Captured by Extraction Wells
Particle Pathlines, Not Captured by Extraction Wells
Extraction Wells
Injection Wells
Model Domain

Color Codes for Horizontal Hydraulic Conductivity (ft/day)

0 - 0.001	10 - 50
0.001 - 0.01	50 - 100
0.01 - 0.1	100 - 300
0.1 - 1	300 - 500
1 - 10	500 - 1000

N

5000 Feet 5000 Feet

J:\Montrose_GW_Remedy\Plots\Pathline_Compare_Jun04.apr

Figure 4-22
Comparison of Particle Tracking Results
Showing Containment in the MBFC
Revised Baseline Calibration and
Containment Failure Scenarios in the MBFC

4.2.3 Predictions of Plume Containment for the Revised Baseline Calibration and Failure Scenarios

Figure 4-22 shows the comparison of particle tracking results for the evaluation of containment in the MBFC for the baseline calibration and failure scenarios. As mentioned above, complete plume containment by the JGWFS wellfield was achieved in the Gage Aquifer by all calibrated solutions including the baseline calibration and four failure scenarios.

The revised baseline calibration predicted almost complete containment of the chlorobenzene plume in the MBFC by the JGWFS wellfield. Both sets of failure scenarios (i.e., Versions 1 and 2), on the other hand, were consistent in predicting similar failure mechanisms for plume containment in the MBFC in spite of differences in the calibration methodology used for these scenarios. Based on these results, it appears that contaminated groundwater could potentially escape containment at the toe of the chlorobenzene plume in the MBFC because of the potential presence of a high hydraulic conductivity zone at this location. Another pathway for escape of contaminated groundwater could exist in the western portion of the chlorobenzene plume in the MBFC (see baseline calibration and 90 percent failure Version 2), also due to the potential presence of a higher hydraulic conductivity zone.

The results of these predictive simulations indicate a significant difference between the performance of the remedial wellfield, with regard to plume containment in the MBFC, for the baseline calibration and failure scenarios. Consequently, there is a considerable uncertainty associated with the ability of the model to predict plume containment in this unit. This uncertainty can be reduced if additional site-specific data could be obtained for the parameters that have the highest contribution to this uncertainty. The discussion of these parameters is presented below.

4.2.4 Model Uncertainty for Predictions of Plume Containment

As discussed above, the qualitative comparison of calibrated model parameters indicated significant differences in values of horizontal hydraulic conductivity at the toe of the chlorobenzene plume in the MBFC, and porosity of the MBFC. In addition to the qualitative comparison of model parameters, the sensitivity of predictive targets to each model parameter was calculated as part of the predictive calibration process. Similar to the uncertainty analysis for the predictions of mass and volume reduction, the estimated sensitivities were used to assess the contribution of each parameter to the predictive uncertainty of the model with regard to plume containment in the MBFC and the Gage Aquifer. A detailed discussion of the methodology used for these estimates is presented in Appendix A and discussed in Moore and Doherty (2004). The estimated contributions of model parameters to the predictive uncertainty pertaining to plume containment are presented in Table 4-4.

TABLE 4-4
Contributions of Model Parameters to Predictive Uncertainty Pertaining to Plume Containment

Model Parameter	Contribution to Uncertainty (i.e., predictive error variance) for Plume Containment Targets (weighted count^2)	
	MBFC	Gage Aquifer
Total	24,000,000	210,000
Porosity	450,000	34
Dispersivity	2400	60
Chlorobenzene source concentration	78,000	240
Chlorobenzene adsorption constant	230,000	130
p-CBSA source concentration	2300	110
Riverbed conductance	1700	9.9
GHB heads	6,000,000	37,000
Recharge	1,800,000	37
K_h in UBF	140,000	390
K_h in MBFB	46,000	590
K_h in MBFM	270	250
K_h in MBFC	15,000,000	430
K_h in LBF	700,000	16,000
K_h in Gage Aquifer	9800	150,000.
K_h in GLA	1200	8100
K_h in Lynwood Aquifer	1700	61
K_v in UBF	130,000	360
K_v in MBFB	6,100	330

NOTE: The contribution to uncertainty is expressed in units of variance for "weighted count of particles", which was used as a surrogate for capture (See Appendix A) – (weighted count)^2. Larger numbers indicate a greater contribution.

Based on these results, the predictive uncertainty of the model with regard to plume containment in the MBFC is significantly higher than that for the Gage Aquifer, which is consistent with the inability of PEST to identify a combination of parameters that would result in the plume containment failure in the Gage Aquifer. Horizontal hydraulic conductivity of the MBFC has, by far, the highest contribution to the uncertainty associated with containing the plume in this layer. GHB conditions and recharge are the second and third most important factors contributing to this uncertainty.

Based on the above analyses including both (1) qualitative evaluation of calibration parameter distributions, and (2) estimates of contribution to uncertainty by each parameter, it appears that additional field data on hydraulic conductivity in the MBFC at the toe of the chlorobenzene plume could result in significant reduction of predictive uncertainty of the model with regard to the plume containment in the MBFC.

Other parameters that contribute to this uncertainty, such as boundary conditions and recharge, are more difficult to measure/determine based on field data. In addition, the contribution of these parameters to the predictive uncertainty is significantly lower than that of hydraulic conductivity in the MBFC. Consequently, the collection of additional data pertaining to these parameters is not recommended at this time. The contribution of these parameters to predictive uncertainty associated with plume containment in the MBFC will be taken into consideration by the RD and performance monitoring program.

4.2.5 Simulation of Failure Scenario with Reduced Hydraulic Conductivity

This section discusses the results of the additional containment failure scenario, which was developed based on the comment from reviewers pertaining to the upper limit of horizontal hydraulic conductivity in the MBFC used in the process of predictive calibration. The essence of the comment was that the interpretation of several aquifer tests presented in the Montrose Remedial Investigation Report (Montrose and EPA, 1998) for the MBFC overestimates hydraulic conductivity for this unit, and the upper limit of 780 ft/day used for the calibration of the hydraulic conductivity range based on these aquifer tests is also too high. Because the failure of the remedial wellfield to contain the plume predicted by the model under the scenarios described above appears to be caused primarily by these high (i.e., equal to the upper limit) values of hydraulic conductivity at the toe of the plume, additional simulations were performed to assess if containment failure could occur with lower values of hydraulic conductivities (i.e., to assess predictive uncertainty of the model with regard to plume containment in the MBFC given the lower value of the upper limit of horizontal hydraulic conductivity in this unit).

An additional scenario was generated using the same calibration and predictive targets and target weights as the 90 percent failure scenario, with respect to plume containment in the MBFC Version 2 (see Section 4.2.2.2). The upper limit of hydraulic conductivity, however, was decreased for this scenario by 50 percent to 340 ft/day. Note, that the upper limit of hydraulic conductivity was decreased only for the purposes of these simulations. Additional aquifer/pilot testing and evaluation/reinterpretation of the existing remedial investigation aquifer test data will be required to verify if such decrease is appropriate for future modeling analysis and remedial design.

The scatter plots of measured and simulated water levels for this scenario is shown in Figure 4-23. The calibrated historical chlorobenzene plumes are shown in Figure 4-24. Based on these results, the model appears to be calibrated reasonably well.

The distribution of calibrated horizontal hydraulic conductivity for the MBFB, MBFC, LBF, and the Gage Aquifer is shown in Figure 4-25. As with other containment failure scenarios, PEST assigned the upper limit value of hydraulic conductivity in the MBFC to the area at the toe of the chlorobenzene plume. However, for this simulation, this value was 50 percent smaller (i.e., 340 ft/day) than for other failure scenarios.

The calibrated transport parameters for this scenario are presented in Table 4-5. The values of transport parameters are comparable with the revised baseline calibration. The porosity of the MBFC of 0.3 used for this scenario is more realistic than that of 0.4 used by PEST in other failure scenarios. As discussed in Section 4.2.2.3, higher values of porosity were likely used by PEST to compensate for high values of hydraulic conductivity at the toe of the chlorobenzene plume. Because the hydraulic conductivity values used for this scenario were 50 percent lower, the lower porosity of 0.3 was sufficient to meet calibration targets.

FIGURE 4-23
Simulated vs. Measured Water Levels – Containment Failure Scenario with Reduced K

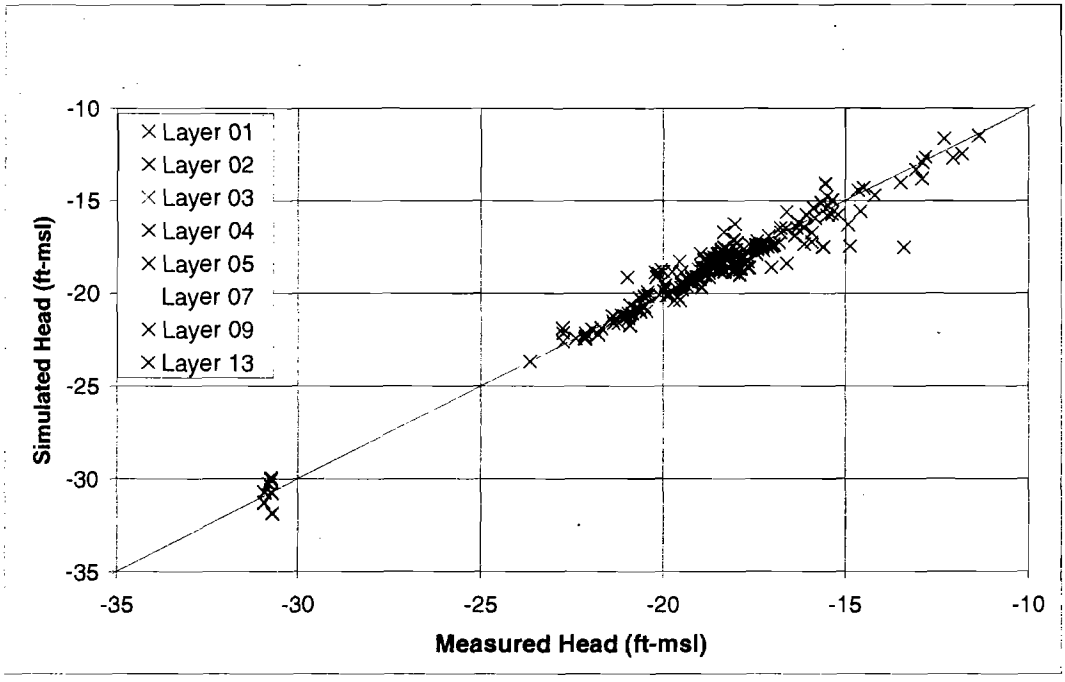
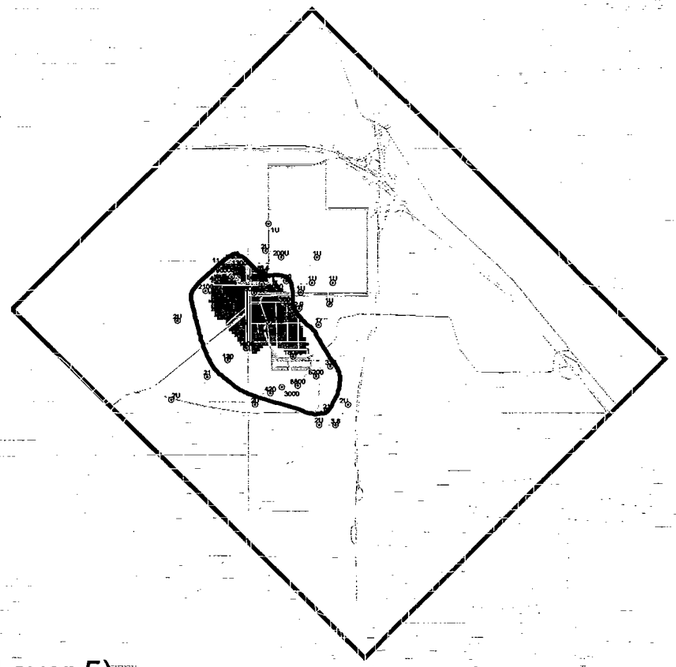


TABLE 4-5
Calibrated Transport Parameters for the Containment Failure Scenario with Reduced K

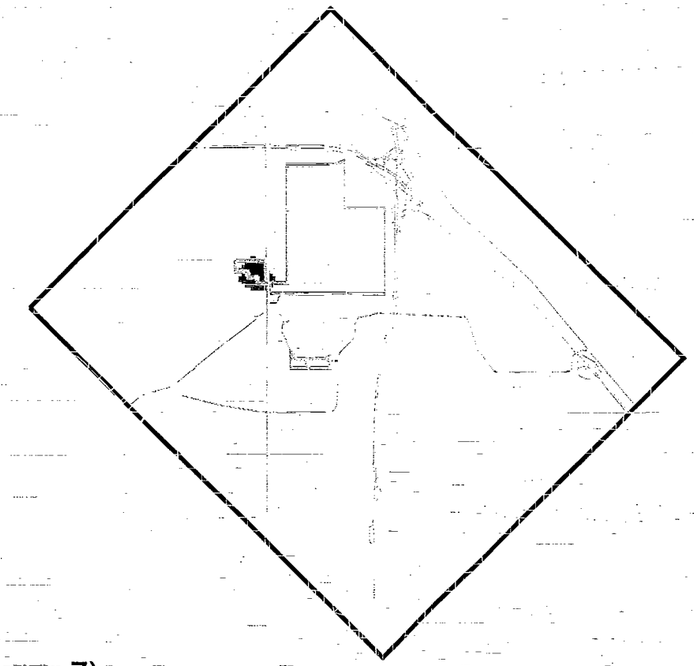
HSU	Porosity	D_L (ft)	D_T/D_L	D_Z/D_L	Chlorobenzene (µg/L)	Chlorobenzene Kd (mL/g)	R (n_{eff} = n)	p-CBSA (µg/L)	Bulk Density (g/cm3)
UBF	0.33	110	0.49	0.00094		0.0053	1.02		1.49
MBFB	0.20	1.6	0.049	0.0026	220000	0.025	1.19	480000	1.49
MBFM	0.26	43			220000	0.017	1.08	480000	1.25
MBFC	0.30	1.2	0.076	0.00018	29000	0.4	3.12	94000	1.59
LBF	0.24	15	0.5	0.00039	9100	0.57	4.61	200000	1.52
Gage	0.13	3.1	0.19	0.00098	9100	0.042	1.49	36000	1.53
GLA	0.29	36	0.5	0.0006		0.6	4.12		1.51
Lynwood	0.28	70	0.52	0.00094		0.052	1.33		1.76

Figure 4-26 shows particle tracking results for the evaluation of containment in the MBFC for this scenario. These results are consistent with other failure scenarios in predicting similar failure mechanisms for plume containment in the MBFC, in spite of the lower hydraulic conductivity at the toe of the plume used for this scenario. Based on these results, it appears that hydraulic conductivity of 340 ft/day at the toe of the chlorobenzene plume could cause contaminated groundwater to escape containment in the MBFC. These results also suggest another pathway for escape of contaminated groundwater in the western portion of the chlorobenzene plume, which also is consistent with previous results (see Section 4.2.3). Similar to other failure scenarios, complete plume containment by the JGWFS wellfield was achieved in the Gage Aquifer under this scenario.

The results of this predictive simulation indicate that considerable uncertainty is associated with the ability of the model to predict plume containment in the MBFC even if the upper limit of hydraulic conductivity for this unit is reduced by 50 percent. This further confirms the need for additional site-specific data pertaining to the hydraulic conductivity in the MBFC.



MBFC (Model Layer 5)



LBF (Model Layer 7)

Gage (Model Layer 9)

Chlorobenzene Concentration (µg/L), January 2004

Interpreted 70 µg/L Chlorobenzene Concentration Contour in MBFC, January 2004

Streets

Site Location

Model Domain

Base Map

Color Codes for Simulated Chlorobenzene Plumes (µg/L)

- 70 - 1000
- 1000 - 5000
- 5000 - 10000
- 10000 - 100000
- 100000 - 1000000

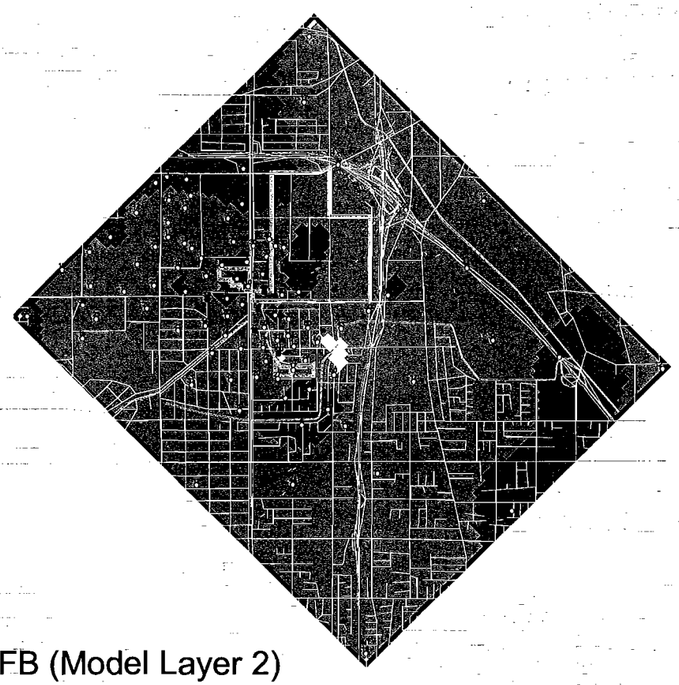
N

5000 Ft 0 5000 Ft

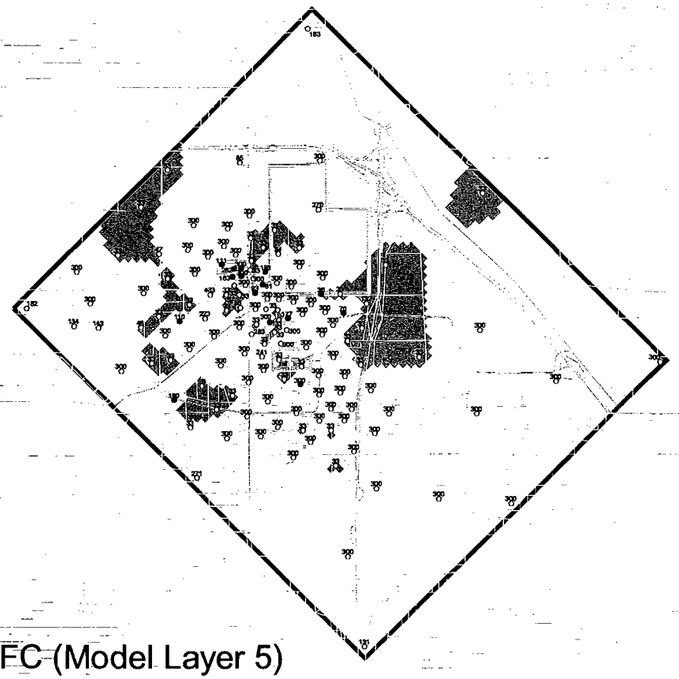
Figure 4-24
Simulated Historical Chlorobenzene Plumes (1945 - 2004)

Containment Failure Scenario
with Reduced Hydraulic Conductivity

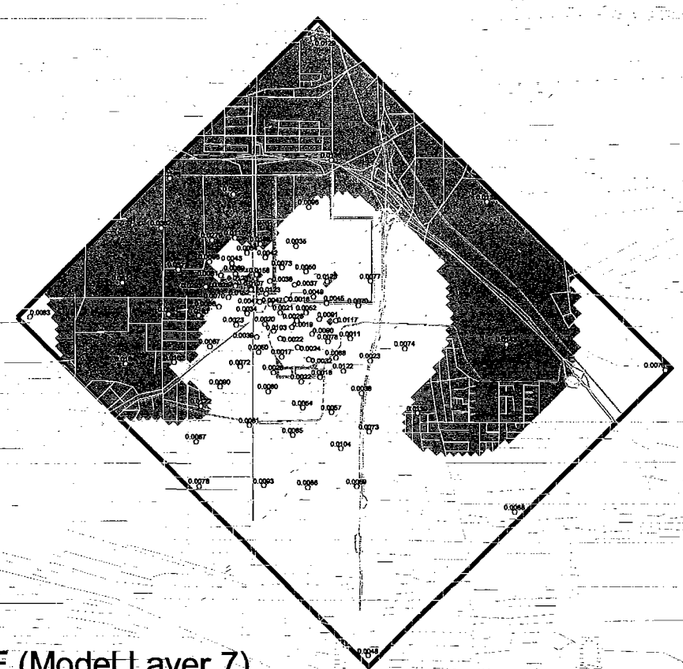
J:\Montrose_GW_Remedy\Plots\Calib_Plume_Maxk.apr



MBFB (Model Layer 2)



MBFC (Model Layer 5)



LBF (Model Layer 7)

Gage (Model Layer 9)

- Pilot Points, Values from Model Calibration
- Pilot Points, Values from Aquifer Tests
- Streets
- Site Location
- Model Domain
- Base Map

Color Codes for Horizontal Hydraulic Conductivity (ft/day)

- 0-0.001
- 0.001 - 0.01
- 0.01 - 0.1
- 0.1 - 1
- 1 - 10
- 10 - 50
- 50 - 100
- 100 - 300
- 300 - 500
- 500 - 1000

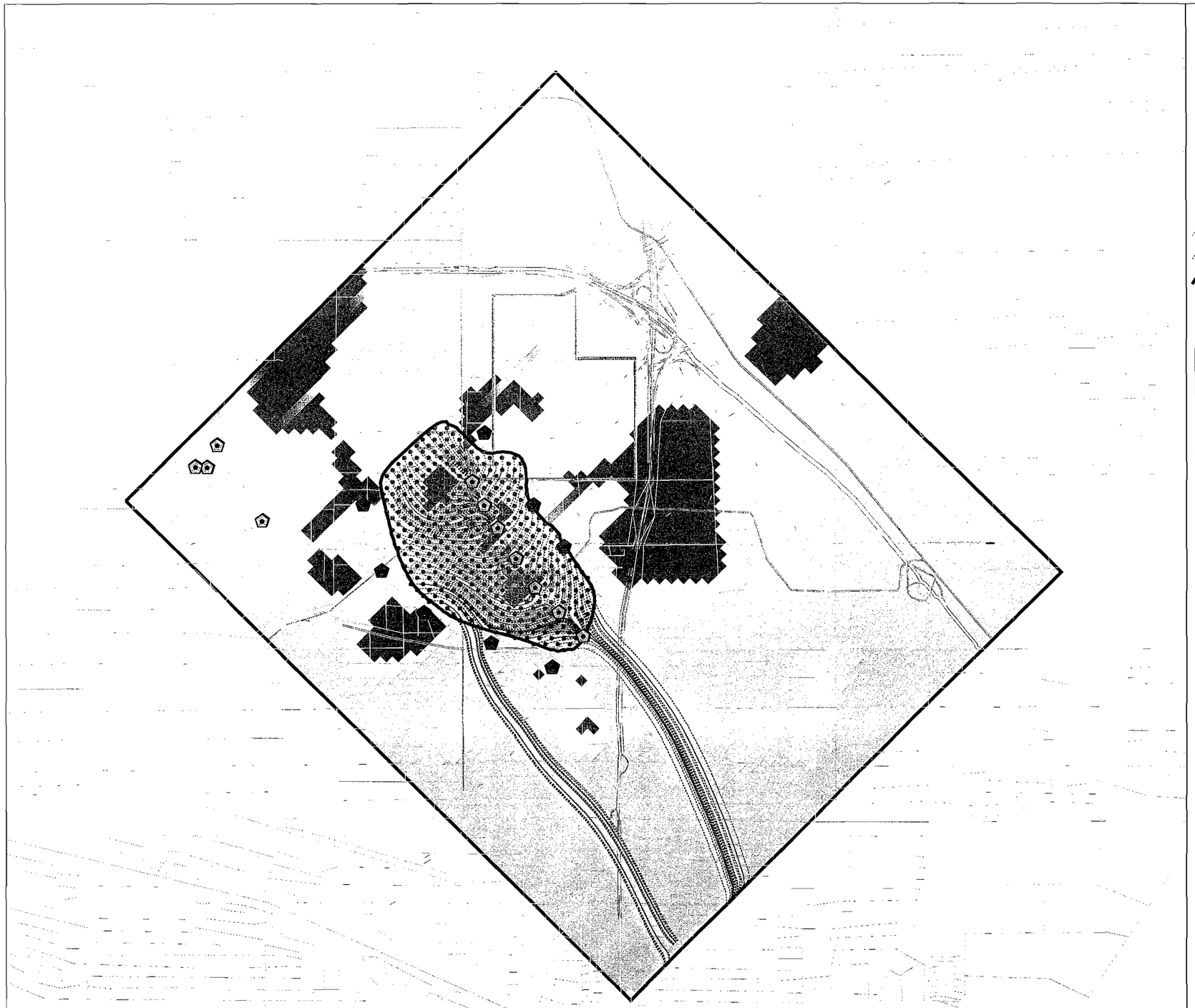
N

5000 Ft 0 5000 Ft

Figure 4-25
Calibrated Horizontal Hydraulic Conductivity

Containment Failure Scenario
with Reduced Hydraulic Conductivity

J:\Montrose_GW_Remedy\Plots\Calib_K_maxK.apr



- Injection Wells
- Extraction Wells
- Particle Starting Locations, Captured by Extraction Wells
- Particle Starting Locations, Not Captured by Extraction Wells
- Particle Pathlines, Captured by Extraction Wells
- Particle Pathlines, Not Captured by Extraction Wells
- Interpreted 70 µg/L Chlorobenzene Contour, January 2004

- Streets
- Site Location
- Model Domain
- Base Map

Color Codes for Horizontal Hydraulic Conductivity (ft/day)

- 0-0.001
- 0.001 - 0.01
- 0.01 - 0.1
- 0.1 - 1
- 1 - 10
- 10 - 50
- 50 - 100
- 100 - 300
- 300 - 500
- 500 - 1000

N

2500 Ft 0 2500 Ft

Figure 4-26
Particle Tracking Results
Showing Containment in the MBFC

Containment Failure Scenario
with Reduced Hydraulic Conductivity

J:\Montrose_GW_Remedy\Plots\Calib_K_maxK.apr

5. Conclusions and Recommendations

Based on the results of this data gap analysis, the following conclusions and recommendations have been drawn:

1. Considerable uncertainty is associated with the ability of the model to predict the performance of the remedial wellfield with regard to contaminant mass and volume reduction rates, and plume containment in the MBFC. The uncertainty associated with plume containment in this unit is significant even if the upper limit of the range of hydraulic conductivity, allowed for model calibration, in the MBFC is reduced by 50 percent.

2. The predictive uncertainty of the model with regard to contaminant mass and volume reduction and plume containment in the Gage Aquifer is low (i.e., the certainty of plume containment and cleanup in the Gage Aquifer as predicted by the model is high).

3. Predictive uncertainty of the model with respect to mass and volume reduction rates in the MBFC is most affected by Kd, which has the highest contribution to this uncertainty. Because of the significant variability of this parameter in natural systems such as that at the Site, and because the field experiments required to quantify this parameter are complicated, costly, time-consuming, and ordinarily ineffective, uncertainty associated with Kd cannot be appreciably reduced. Because of the highest contribution from Kd, the predictive uncertainty of the model with respect to mass and volume reduction in the MBFC is considered to be mostly irreducible and should be accounted for by the RD and performance monitoring program.

4. Predictive uncertainty of the model with regard to plume containment in the MBFC is most affected by horizontal hydraulic conductivity of the MBFC. Additional field data on hydraulic conductivity in the MBFC could result in significant reduction of predictive uncertainty of the model with regard to the plume containment in the MBFC. Additional hydraulic conductivity data in the MBFC may also result in some reduction of predictive uncertainty pertaining to the rates of volume reduction in the MBFC.

5. All containment failure scenarios predicted similar failure mechanisms for plume containment in the MBFC. Based on these results, it appears that contaminated groundwater could potentially escape containment at the toe of the chlorobenzene plume in the MBFC because of the potential presence of a high hydraulic conductivity zone at this location. Another pathway for escape of contaminated groundwater could exist in the western portion of the chlorobenzene plume in the MBFC, also due to the potential presence of a higher hydraulic conductivity zone. Based on these results, additional hydraulic conductivity data should be collected in these areas to reduce predictive uncertainty of the model with regard to plume containment.

6. The initial model runs performed in the process of the baseline calibration indicated the potential leakage of chlorobenzene from the high-concentration plume in the MBFC in the vicinity of well BF-21 into the area west of the currently defined Gage Aquifer plume. Based on these results, the potential leakage from the MBFC west of the currently defined plume in the Gage Aquifer needs to be assessed, because this area is located downgradient of nondetect well G-16, which currently delineates the western edge of the chlorobenzene plume in the Gage Aquifer.

7. The extent of the chlorobenzene plume in the Gage Aquifer needs to be delineated south of well G-19.

8. The use of PEST linked with the MODFLOW2000, MT3DMS, and MODPATH is appropriate for the development and calibration of the RD model and design of the remedial wellfield(s) for the site.

In summary, the certainty of model predictions with regard to containing the plume and achieving clean up rates in accordance with the ROD criteria is low for the MBFC and relatively high for the Gage Aquifer, based on the results of both mass and volume reduction and containment failure scenarios. Kd contributes the most uncertainty regarding cleanup rates, but in practice, cannot be reliably measured in the field. Hydraulic conductivity in the MBFC is the primary source of uncertainty with respect to plume containment, as well as a secondary contributor with respect to cleanup rates. Consequently, additional hydraulic conductivity measurements in the MBFC should be collected during the pilot test program. The impact of parameters which contribute to predictive uncertainty but can not be assessed in the field, such as Kd and boundary conditions, will be further assessed during the RD modeling and considered in the remedial wellfield design and performance monitoring program. In addition, if data collected during the RD results in different assumptions regarding the distribution of model parameters, it is possible that other failure mechanisms than those discussed in this report can be generated. While Kd and hydraulic conductivity are expected to contribute the most uncertainty to clean up and containment predictions, respectively, during RD modeling, it is possible that additional parameters with high contributions to uncertainty will be identified during RD model development. If so, they may need to be assessed as part of the second phase of pilot testing, and/or considered in the design of the remedial wellfield and performance monitoring program.

6. References

Bedient, P. B., H. S. Rifai, and C. J. Newell. 1999. Ground Water Contamination, Prentice Hall, New Jersey.

CH2M HILL. 1998. Final Joint Groundwater Feasibility Study for the Montrose and Del Amo Sites. May 18.

CH2M HILL. 2003. Work Plan for Development of Groundwater Model for Remedial Design, Montrose/Del Amo Dual Site. September.

Doherty, John. 2002. Groundwater Model Calibration using Pilot Points and Regularization. *Ground Water*. Vol. 41 (2): 170-177.

Doherty, John and John M. Johnston. 2003. Methodologies for Calibration and Predictive Analysis of a Watershed Model, *J. American Water Resources Association*, 39(2):251-265.

Moore, Catherine and John Doherty. 2004. The Role of the Calibration Process in Reducing Model Predictive Error.

Fu, G., A. T. Kan, and M. B. Tomson. 1994. Adsorption and desorption irreversibility of polycyclic aromatic hydrocarbons in surface sediment. I. characterization of desorption hysteresis, *Environ. Toxicol. Chem.*, 13(10), 1559-1567.

Hargis + Associates. 2003. Preliminary Groundwater Modeling, Montrose Site, Torrance, California. May 30.

Hargis + Associates. 2004. Baseline Groundwater Sampling Results Report, Montrose Site, Torrance, California. April 21.

Montrose Chemical Corporation and United States Environmental Protection Agency (EPA). 1998. Final Remedial Investigation Report for the Montrose Superfund Site. May 18.

Kan, A. T., G. Fu, M. A. Hunter, and M. B. Tomson. 1997. Irreversible adsorption of naphthalene and tetrachlorobiphenyl to Lula and surrogate sediments, *Environ. Sci. Technol.*, 31, 2176-2185.

Pignatello, J. P. and B. Xing. 1995. Mechanisms of slow sorption of organic chemicals to natural particles, *Environ. Sci. Technol.*, 29, 1-10.

S.S. Papadopulos & Associates, Inc. 1996. MT3D: A Modular Three-Dimensional Transport Model, Version 1.5, Documentation and User's Guide. March 15, 1992, and 1996.

United States Environmental Protection Agency (EPA). 1999. Record of Decision for Dual-Site Groundwater Operable Unit Montrose Chemical and Del Amo Superfund Sites. March.

United States Geological Survey (USGS). 1994. User's Guide for MODPATH/MODPATH-PLOT, Version 3. Open-File Report 94-464. September.

United States Geological Survey (USGS). 1988. A Modular Three-Dimensional Finite-Difference Groundwater Flow Model. Open-File Report 83-875.

United States Geological Survey (USGS). 2000. MODFLOW-2000, The U.S. Geological Survey Modular Ground-Water Model. Open-File Report 00-92.

Zheng, C. and P. P. Wang. 1999. MT3DMS: A Modular Three-Dimensional Multispecies Transport Model. November.

Appendix A

The Role of the Calibration Process in Reducing Model Predictive Error (Moore and Doherty, 2004)

The Role of the Calibration Process in Reducing Model Predictive Error

Catherine Moore and John Doherty

School of Engineering, University of Queensland, Australia.

s4024009@student.uq.edu.au, jdoherty@gil.com.au.

Abstract

An equation is derived through which the variance of predictive error of a calibrated model can be calculated. This equation has two terms. The first term represents the contribution to predictive error variance that results from an inability of the calibration process to capture all of the parameterization detail that is necessary for the making of an accurate prediction. If a model is "uncalibrated", with parameter values being supplied solely through "outside information", this is the only term required. The second term represents the contribution to predictive error variance arising from measurement noise. In an overdetermined system, such as may be obtained through "parameter lumping" (for example through the introduction of a spatial zonation scheme), this is the only term required. It is shown, however, that parameter lumping is a form of "implicit regularization", and that ignoring the implied first term of the predictive error variance equation can potentially lead to underestimation of predictive error variance.

A model's role as a predictor of environmental behavior can be enhanced if it is calibrated in such a way as to reduce the variance of those predictions which it is required to make. It is shown that in some circumstances this can be accomplished through

"overfitting" against historical field data. It can also be accomplished by giving greater weight to those measurements which carry greatest information content with respect to a required prediction. This raises the specter that a single "calibrated model" may not perform as well in environmental management as a number of models, each of which is calibrated in such a way as to optimize its accuracy in making a prediction of a specific type.

Introduction

Models are often used to make predictions of environmental behavior. Used in that role they support environmental management. A vital aspect of the model construction process is the calibration phase. During this phase model parameters are adjusted until the model's replication of historical field measurements is judged to be "reasonably good". It is then assumed that this constitutes sufficient justification to use the model to make predictions, and that those predictions will also be "reasonably good".

Unfortunately even predictions made by a model that matches historical data perfectly may be considerably in error (Moore and Doherty, 2004a). In fact, a model's predictive uncertainty will only be reduced by calibration if the information content of the calibration dataset is able to constrain those parameters that have a significant bearing on that prediction. Thus, for example, if a model is built to make predictions of contaminant transport, then its performance in this regard will be better served if it is calibrated against historical contaminant concentration data, than if it is calibrated solely against groundwater heads. (see, for example, Harvey and Gorelick, 1995; Poeter and Belcher 1991; Frederick and Doherty, 2003; Franssen et al., 2003; and Feyen et al., 2003). For the

same reason, Tiedeman et al. (2003) assert that model predictive uncertainties may not necessarily decrease with calibration to a larger dataset.

Model predictive uncertainty arises from a number of sources. In the present paper inadequacies in model equations, or in the numerical implementation of those equations, will be neglected, and only those pertaining to its parameterization will be considered.

Information on model parameters comes from two sources. The first of these sources is "prior information" on parameter values originating from direct measurement of system properties, or from what can be inferred about those properties through knowledge of the materials of which the system under study is composed. Such information is often vague, or at best point-based, so is often best expressed in stochastic terms; for models such as groundwater models where parameterization is spatially-based, geostatistical characterizations of hydraulic properties are often thus employed. The second source of knowledge on model parameters arises from historical measurements of system state. This is a more indirect form of knowledge of system properties, the information content of which is "tapped" during the model calibration process. Unfortunately, this information is often contaminated by noise; thus parameter values inferred from such data have a stochastic component that originates in the uncertainty associated with field measurements, as well as in so-called "structural noise" - a term used to describe various unavoidable forms of model inadequacy. Another problem with inferring parameter values from field measurements is that such measurements, like direct measurements of system properties, are often sporadic in both space and time. Hence there is an upper limit to the level of parameterization detail that can be inferred from them; see Backus and

Gilbert (1969), Menke (1984), Kitanidis (1997), Guadagnini and Neuman (1999) and Gorokhovski (1996) to mention just a few discussions of this important subject.

Bayes theorem provides the means to assimilate these two sources of information on model parameters into a "posterior parameter distribution" that reflects both the constraining effects of the calibration process, and knowledge of parameter values that originates from outside of this process. Model predictive probabilities can then be evaluated using the relationships between model parameters and model predictive outputs encapsulated in the model. If necessary, such an analysis can include geostatistical characterizations of prior parameter uncertainty (see for example Woodbury and Ulrich (2000), Woodbury and Rubin (2000) and papers cited therein). A non-Bayesian, but nevertheless effective, means of constraining geostatistical characterizations of parameter spatial variability such that historical measurements of system state are respected by the model, involve the "bending" or "warping" of stochastic seed fields, thus forcing these fields to satisfy calibration constraints. For examples of this methodology see RamaRao et al. (1995), LaVenue et al. (1995), Gómez-Hernándezet al. (2003) and Doherty (2003).

An alternative methodology for analyzing model predictive uncertainty is presented in this paper. Use of this methodology is based on the premise that a model has been "calibrated" against a set of field measurements as a precursor to its deployment for making predictions of future system behavior, this being the most common strategy for using models in environmental management. It is further assumed that calibration takes place as an underdetermined inverse problem. This strategy allows a model to employ a level of complexity that is sufficient to represent all processes on which a prediction of

87 interest depends. While parameters pertaining to that complexity may not be uniquely
88 estimated, it is demonstrated below that full characterization of predictive error variance
89 requires that this complexity be represented. The theory is then extended to
90 overdetermined parameter estimation (applicable, for example, where a model domain is
91 subdivided into a small number of zones of piecewise parameter constancy in accordance
92 with the principle of parsimony). It is demonstrated that estimates of model predictive
93 error variance made as an adjunct to model calibration based on this principal can be
94 seriously flawed unless the effects of such system simplification are taken into account in
95 making these estimates.

96 Solution of an underdetermined inverse problem is possible only if some regularization
97 strategy is employed. The use of regularized inversion in the context of groundwater
98 model calibration has been discussed by a number of authors, including Vasco et al.
99 (1997), Clemo et al. (2003), Doherty (2003) and Moore and Doherty (2004a), (2004b).
100 Software such as PEST (Doherty; 2004) is freely available for its implementation.
101 Through the use or regularized inversion, simplifications in parameterization necessary
102 for the achievement of numerical stability of the inverse problem are undertaken by the
103 parameter estimation process itself, rather than through manual simplification as a
104 precursor to that process. This allows maximum information content to be extracted from
105 a given calibration dataset.

106 The discussion below begins by exploring model predictive uncertainty analysis in
107 contexts where parameterization is unassisted by calibration. After a brief discussion of
108 regularized inversion, exploration of predictive uncertainty analysis is extended to

accommodate the imposition of calibration constraints on parameter values. Through an analysis of the equations so derived, some important points regarding the role of model calibration in reducing (or not reducing) model predictive uncertainty are discussed. With these points in mind the discussion then turns to how the model calibration process can be made to better serve the model predictive process, particularly with regard to the assignment of "measurement weights" to elements of the calibration dataset. Finally some of the concepts developed in the analysis are applied to a synthetic case to demonstrate their use.

It must be pointed out that the equations derived below are based on an assumption of model linearity. Most models, of course, are nonlinear; hence these equations will be only approximations in many cases. Nevertheless they are useful for the contribution that they make to our understanding of the calibration process. Furthermore, it is hoped that their use can extend farther than this, to a semi-quantitative analysis of calibration outcomes. Where such an analysis seeks ordering relationships rather than absolutes (for example in determining the relative contribution to uncertainty made by different parameter groups, or the relative reduction in uncertainty that can be accrued through acquisition of different types of supplementary data), conclusions drawn through application of these equations are likely to be quite robust, notwithstanding the nonlinear nature of a model to which they may be applied.

Theory

Linear Predictive Uncertainty Analysis for an Uncalibrated Model

Suppose that the ($m\times1$) vector **p** contains the values of parameters used by a model. Unless these values are accurately known at all places within a model domain, they must be described in probabilistic terms. Let the covariance matrix of **p** be denoted as C(**p**). Let s (a scalar) designate a prediction made by the model; let the sensitivities of this prediction to model parameters be represented by the vector **y**. Then s is calculable using the relationship:-

$$s = \mathbf{y}^t\mathbf{p} \qquad (1)$$

where the "t" superscript denotes the transpose operation. (Note that in this, and subsequent equations, parameter and prediction offsets are ignored for the sake of simplicity. Thus **p** is to be considered as parameter perturbations from some known average value; this makes no difference to the equations and concepts derived below.) Through basic matrix manipulation it is easily shown that the variance of s (i.e. σ^2_s) is given by:-

$$\sigma^2_s = \mathbf{y}^t C(\mathbf{p})\mathbf{y} \qquad (2)$$

Equation 2 was used by El Harrouni et al. (1997) in calculating output uncertainties for the DRBEM boundary element groundwater model on the basis of spatially correlated model parameters.

Regularized Inversion

Let the (assumed linear) relationship between the m model parameters $\mathbf{p}$ and n model outputs $\mathbf{o}$ be represented by the matrix equation:

$$\mathbf{X}\mathbf{p} = \mathbf{o} \tag{3}$$

where $\mathbf{X}$ is the model "sensitivity" or "Jacobian" matrix. Let $\mathbf{h}$ be a vector of field measurements corresponding to the model output vector $\mathbf{o}$. $\mathbf{h}$ is expressible as:-

$$\mathbf{h} = \mathbf{X}\mathbf{p} + \boldsymbol{\varepsilon} \tag{4}$$

where $\mathbf{p}$ represents the "true" parameters of the model (which we will never know), and $\boldsymbol{\varepsilon}$ represents measurement and structural noise associated with $\mathbf{h}$. Let the covariance of this noise be represented by the $n \times n$ matrix $C(\boldsymbol{\varepsilon})$; for better or for worse, this is normally assumed to be a diagonal matrix.

Let the extent of model-to-measurement misfit be represented by an objective function Φ defined as:-

$$\Phi = (\mathbf{X}\mathbf{p} - \mathbf{h})^t\mathbf{Q}(\mathbf{X}\mathbf{p} - \mathbf{h}) \tag{5}$$

where $\mathbf{Q}$ is a positive definite "cofactor matrix". This is normally chosen to be proportional to the inverse of $C(\boldsymbol{\varepsilon})$, that is:-

$$C(\boldsymbol{\varepsilon}) = \sigma_h^2\mathbf{Q}^{-1} \tag{6}$$

where the constant of proportionality, σ_h^2 (the so-called "reference variance"), can be estimated through the calibration process as:-

$$\sigma_h^2 = \Phi/(n-m) \quad (7)$$

where Φ is the objective function corresponding to an acceptable level of model-to-measurement fit.

Φ of equation 5 is minimized when:-

$$\mathbf{X^tQX\underline{p} = X^tQh} \quad (8)$$

If the matrix $\mathbf{X^tQX}$ is not of full rank and the inverse problem is thus underdetermined, there is no unique solution to equation 8.

In highly parameterized contexts minimization of Φ can lead to "over-fitting". Hence it is often better to look for a parameter set $\underline{\mathbf{p}}$ that results in an objective function Φ_n which is somewhat higher than minimal. That is, we seek a $\underline{\mathbf{p}}$ which satisfies the equation:-

$$(\mathbf{X\underline{p} - h})^t\mathbf{Q}(\mathbf{X\underline{p} - h}) = \Phi_n \quad (9)$$

Φ_n should be set at a value that reflects the measurement and structural noise content of the observation dataset on which calibration is based. Whether or not $\mathbf{X^tQX}$ is of full rank, solution of equation 9 is nonunique if Φ_n is greater than the minimized objective function. Thus it must be solved through some kind of regularized inversion process. The present discussion focuses on "truncated singular value decomposition" as a regularization mechanism. However the conclusions, and many of the equations, derived below are just as applicable to other regularization methods such as "constrained minimization regularization", otherwise known as "Tikhonov regularization".

185 Singular value decomposition (SVD) can be used to determine the eigenvalues and
186 eigenvectors of $\mathbf{X}^t\mathbf{QX}$, whether or not this matrix is of full rank. Thus:-

187 $$\mathbf{X}^t\mathbf{QX} = \mathbf{VEV}^t \qquad (10)$$

188 where $\mathbf{V}$ is the matrix of eigenvectors of $\mathbf{X}^t\mathbf{QX}$ and $\mathbf{E}$ is a diagonal matrix of eigenvalues
189 of $\mathbf{X}^t\mathbf{QX}$. Where $\mathbf{X}^t\mathbf{QX}$ has less than full rank some of the eigenvalues of $\mathbf{E}$ are zero; in
190 fact $\mathbf{E}$ has as many zero-valued eigenvalues as the rank-deficiency of $\mathbf{X}^t\mathbf{QX}$. Because
191 $\mathbf{X}^t\mathbf{QX}$ is positive semi-definite, its eigenvalues are real, and its eigenvectors are
192 orthogonal. Thus:-

193 $$\mathbf{V}^t = \mathbf{V}^{-1} \qquad (11)$$

194 Let $\mathbf{V}$ be characterized as:

195 $$\mathbf{V} = [\mathbf{V}_1\ \mathbf{V}_2] \qquad (12)$$

196 where $\mathbf{V}_1$ contains eigenvectors corresponding to the k largest eigenvalues of $\mathbf{X}^t\mathbf{QX}$, and
197 $\mathbf{V}_2$ contains the remaining eigenvectors, including those whose eigenvalues are zero. In
198 order to obtain a unique solution for $\underline{\mathbf{p}}$ at an acceptable level of model-to-measurement
199 misfit, let all eigenvectors after the k'th be assigned a value of zero (hence the term
200 "truncated" in the description of this regularization methodology). Pre-multiplication of
201 equation 8 by $\mathbf{V}^t_1\mathbf{V}_1\mathbf{E}^{-1}{}_1\mathbf{V}^t_1$ then results in:-

202 $$\mathbf{V}^t_1\underline{\mathbf{p}} = \mathbf{E}^{-1}{}_1\mathbf{V}^t_1\mathbf{X}^t\mathbf{Qh} \qquad (13)$$

203 In equation 13 $\mathbf{E}_1$ is the diagonal matrix of non-zero eigenvalues of $\mathbf{X}^t\mathbf{QX}$. The elements
204 of the vector $\mathbf{V}^t_1\underline{\mathbf{p}}$ are the inner product of a parameter solution vector with each of the

eigenvectors contained in $\mathbf{V}_1$. Thus equation 13 solves for the projection of solutions of equation 8 onto the subspace of parameter space spanned by the eigenvectors contained in $\mathbf{V}_1$. Because $\mathbf{E}_1$ has no diagonal elements equal or close to zero (which is ensured if k is selected low enough), a stable solution to the regularized inversion problem has been obtained:-

$$\underline{\mathbf{p}} = \mathbf{V}_1\mathbf{E}^{-1}{}_1\mathbf{V}^t{}_1\mathbf{X}^t\mathbf{h} \qquad (14)$$

Thus the m-dimensional inverse problem has been transformed into a k-dimensional inverse problem confined to the subspace of parameter space spanned by $\mathbf{V}_1$. Furthermore, by choosing k appropriately, a good, but not excessively good, fit can be obtained between model outputs and field data. Normally k is chosen such that (9) is approximately obeyed.

In general, eigenvectors of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ (ie. columns of $\mathbf{V}$) corresponding to high eigenvalues show low spatial variability within the model domain, whereas those corresponding to low eigenvalues tend to show high variability (Moore and Doherty 2004b, Wiggins et al. 1976). Thus the truncated SVD solution process tends to select smooth solutions to the inverse problem, this reflecting the inherent incapacity of a calibration dataset to furnish fine system detail in most modeling contexts.

Linear Predictive Uncertainty Analysis for a Calibrated Model

If equation 4 is substituted into equation 14, we obtain:-

$$\underline{\mathbf{p}} = \mathbf{V}_1\mathbf{E}^{-1}{}_1\mathbf{V}^t{}_1\mathbf{X}^t\mathbf{Q}(\mathbf{X}\mathbf{p} + \boldsymbol{\varepsilon}) \qquad (15)$$

Expanding terms in this equation and substituting (10), it becomes:-

$$\underline{\mathbf{p}} = \mathbf{V}_1\mathbf{V}^t_1\mathbf{p} + \mathbf{V}_1\,\mathbf{E}^{-1}{}_1\mathbf{V}^t_1\mathbf{X}^t\mathbf{Q}\boldsymbol{\varepsilon} \qquad (16a)$$

that is:-

$$\underline{\mathbf{p}} = \mathbf{R}\mathbf{p} + \mathbf{G}\boldsymbol{\varepsilon} \qquad (16b)$$

where **R**, the so-called "resolution matrix", describes the relationship between estimated parameters and "true" parameters. The difference between true and estimated parameters is given by:-

$$\mathbf{p} - \underline{\mathbf{p}} = (\mathbf{I} - \mathbf{R})\mathbf{p} - \mathbf{G}\boldsymbol{\varepsilon} \qquad (17)$$

Equation 17 expresses the "parameterization wrongness" of a calibrated model; unfortunately this cannot be calculated because **p** is unknown. However its expected value (i.e. $E(\mathbf{p} - \underline{\mathbf{p}})$) is equal to the expected value of **p** (i.e. $E(\mathbf{p})$), assuming that $E(\boldsymbol{\varepsilon})$ is zero. As was stated above, with the elements of **p** defined as parameter perturbations from their (assumed known) average values, $E(\mathbf{p} - \underline{\mathbf{p}})$ is zero.

Let it be assumed that the covariance matrix of **p** (i.e. $C(\mathbf{p})$ as featured in equation 2), and the covariance matrix of measurement and structural noise (i.e. $C(\boldsymbol{\varepsilon})$ as featured in equation 6), are known; let it be further assumed that **p** and $\boldsymbol{\varepsilon}$ are independent. Then $C(\mathbf{p} - \underline{\mathbf{p}})$ is easily calculated from (17) as:-

$$C(\mathbf{p} - \underline{\mathbf{p}}) = (\mathbf{I} - \mathbf{R})C(\mathbf{p})(\mathbf{I} - \mathbf{R})^t + \mathbf{G}C(\boldsymbol{\varepsilon})\mathbf{G}^t \qquad (18)$$

From equations 16 it is apparent that where truncated SVD is used as a regularization device:

$$\mathbf{R} = \mathbf{V}_1\mathbf{V}^t_1 \tag{19}$$

and:-

$$\mathbf{G} = \mathbf{V}_1\,\mathbf{E}^{-1}{}_1\mathbf{V}^t_1\mathbf{X}^t\mathbf{Q} \tag{20}$$

If these relationships plus (6) are now substituted into (18), and use is made of the relationship:-

$$\mathbf{V}\mathbf{V}^t = (\mathbf{V}_1\mathbf{V}_1{}^t + \mathbf{V}_2\mathbf{V}_2{}^t) = \mathbf{I} \tag{21}$$

we obtain:-

$$C(\mathbf{p} - \underline{\mathbf{p}}) = \mathbf{V}_2\mathbf{V}^t_2C(\mathbf{p})\,\mathbf{V}_2\mathbf{V}^t_2 + \sigma_h{}^2\mathbf{V}_1\mathbf{E}_1{}^{-1}\mathbf{V}^t_1 \tag{22}$$

This expression can be made even simpler if the pre-calibration probability distribution $C(\mathbf{p})$ of model parameters is such that they are all independently variable and have the same variance $\sigma_p{}^2$. In this case equation 18 becomes:-

$$C(\mathbf{p} - \underline{\mathbf{p}}) = \sigma_p{}^2\mathbf{V}_2\mathbf{V}^t_2 + \sigma_h{}^2\mathbf{V}_1\mathbf{E}_1{}^{-1}\mathbf{V}^t_1 \tag{23}$$

If a model prediction s is calculated from model parameters $\mathbf{p}$ using equation 1, we can compare the "model-calculated prediction" (i.e. the prediction made on the basis of calibrated parameters $\underline{\mathbf{p}}$, designated herein as $\underline{s}$) with the "true" prediction (i.e. the prediction made on the basis of "true" parameters $\mathbf{p}$, designated herein as s) using the formula:-

$$s - \underline{s} = \mathbf{y}^t(\mathbf{p} - \underline{\mathbf{p}}) \tag{24}$$

Equation 24 expresses the "wrongness" or "error" of a model prediction. Once again, this can never be known. For the same reasons as those already stated with respect to parameter "wrongness", its preferred value is zero. Its variance can be calculated from (24) as:-

$$\sigma^2_{s-\underline{s}} = \mathbf{y}^t\mathbf{C}(\mathbf{p} - \underline{\mathbf{p}})\mathbf{y} \tag{25}$$

Substitution of this relationship into (18), (22) and (23) yields:-

$$\sigma^2_{s-\underline{s}} = \mathbf{y}^t(\mathbf{I} - \mathbf{R})\mathbf{C}(\mathbf{p})(\mathbf{I} - \mathbf{R})^t\mathbf{y} + \mathbf{y}^t\mathbf{G}\mathbf{C}(\varepsilon)\mathbf{G}^t\mathbf{y} \tag{26a}$$

$$\sigma^2_{s-\underline{s}} = \mathbf{y}^t\mathbf{V}_2\mathbf{V}^t_2\mathbf{C}(\mathbf{p})\ \mathbf{V}_2\mathbf{V}^t_2\mathbf{y} + \sigma_h^2\mathbf{y}^t\mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}^t_1\mathbf{y} \tag{26b}$$

$$\sigma^2_{s-\underline{s}} = \sigma_p^2\mathbf{y}^t\mathbf{V}_2\mathbf{V}^t_2\mathbf{y} + \sigma_h^2\mathbf{y}^t\mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}^t_1\mathbf{y} \tag{26c}$$

Equation 26a is perfectly general, and pertains to no regularization method in particular. Thus exactly the same formula can be used to calculate predictive error variance where regularization is achieved through a constrained minimization process. However in that case, equations 19 and 20 are replaced by:-

$$\mathbf{R} = (\mathbf{X}^t\mathbf{Q}\mathbf{X} + \beta^2\mathbf{Z}^t\mathbf{Q}_r\mathbf{Z})^{-1}\mathbf{X}^t\mathbf{Q}\mathbf{X} \tag{27a}$$

$$\mathbf{G} = (\mathbf{X}^t\mathbf{Q}\mathbf{X} + \beta^2\mathbf{Z}^t\mathbf{Q}_r\mathbf{Z})^{-1}\mathbf{X}^t\mathbf{Q} \tag{27b}$$

where $\mathbf{Z}$ is a matrix of regularization constraints on parameter values (assumed to be linear), $\mathbf{Q}_r$ is the regularization covariance matrix, and β^2 is the "regularization weight factor", which can also be considered to be a Lagrange multiplier in the constrained

minimization process; see Doherty (2003) for a description of this type of regularization, and for an example of its use in the groundwater modeling context.

Significance of Equations

Equations 26 are of great importance. Formulation of predictive error variance using these equations has the benefit that the contributions made to this variance by two different aspects of the model parameterization process are made explicit. The second term of equations 26 is the component of model predictive uncertainty that arises from model-to-measurement misfit. In an overdetermined system (where parameters are outnumbered by observations), this is the only source of model predictive uncertainty, for under these conditions the resolution matrix **R** is actually the identity matrix, and the first term of equations 26 vanishes. Model predictive error analysis based on this term has been undertaken by a number of authors in the groundwater modeling context; see, for example, Hill (1989) and Christensen and Hill (1999). Vecchia and Cooley (1987) extended its use to nonlinear models.

The first term of equations 26 accommodates the fact that the calibration process cannot capture all of the hydraulic detail prevailing within a study area. The further removed is the resolution matrix **R** from the identity matrix **I**, the larger is this term. In general, both data scarcity and high data noise content promote "blurry" resolution matrices, and hence loss of system detail in a calibrated model. This can lead to grossly inaccurate model predictions, where these predictions depend on that detail.

For an uncalibrated model the second term of equations 26 is zero and the resolution matrix **R** becomes the null **0** matrix (this can be demonstrated by setting **Q** to zero –

effectively giving all observations a weight of zero in the inversion process). Equation 26a then becomes equation 2, as it should.

The second term of equations 26 also becomes zero when a perfect fit is obtained between model outputs and field measurements because of the absence of any measurement or structural noise (i.e. $C(\boldsymbol{\varepsilon})$ is $\mathbf{0}$). The fact that predictions made by a "perfectly calibrated" model can be substantially in error is readily apparent from an inspection of the $\mathbf{y}^t(\mathbf{I}-\mathbf{R})$ portion of the remaining term. For an underdetermined system $\mathbf{R}$ is rank-deficient; therefore its columns span only a subspace of parameter space. It is thus possible for $\mathbf{y}^t\mathbf{R}$ to be zero; under these circumstances the calibration process does nothing whatsoever to reduce the uncertainty of that particular model prediction, for equation 26a then yields the same results as equation 2.

This analysis can be taken a step further. When model-to-measurement misfit is zero, $\mathbf{R}$ becomes a projection operator. This is easily demonstrated by observing that:-

$$\mathbf{RRp} = \mathbf{R}\underline{\mathbf{p}} \quad (\text{from } \underline{\mathbf{p}} = \mathbf{Rp}) \tag{28a}$$

and

$$\mathbf{R}\underline{\mathbf{p}} = \underline{\mathbf{p}} \tag{28b}$$

and therefore

$$\mathbf{RRp} = \underline{\mathbf{p}} = \mathbf{Rp}, \text{ and thus } \mathbf{RR} = \mathbf{R} \tag{29}$$

Equation 28b follows from the fact that $\mathbf{Xp} = \mathbf{X}\underline{\mathbf{p}}$. Also following from this is the fact that:-

$$\mathbf{XR} = \mathbf{X} \tag{30}$$

from which, with a little basic matrix manipulation, it can be demonstrated that:

if $\mathbf{Xy}$ is $\mathbf{0}$, then $\mathbf{Ry}$ is $\mathbf{0}$ (31)

From (31) it follows that $\mathbf{y}^t\mathbf{R}$ will be zero if $\mathbf{Xy}$ is $\mathbf{0}$ and $\mathbf{R}$ is symmetrical, as it always is when regularized inversion is implemented using truncated SVD. From this it follows that if, for a particular model prediction, the vector of predictive sensitivities (i.e. $\mathbf{y}$) is perpendicular to all observation sensitivity vectors (i.e. to all rows of the $\mathbf{X}$ matrix), then the calibration process does nothing to decrease the uncertainty of this prediction. Thus if a model is calibrated against data types which bear little relation to the types of predictions that a model will be required to make, then there can be no guarantee that the calibration process will reduce the uncertainties of these predictions at all.

Examination of equation 26c allows further insight to be gained into the role of the calibration process in reducing (or not) predictive uncertainty. Consider that a prediction sensitivity vector $\mathbf{y}$ is parallel to an eigenvector of $\mathbf{V}$. That particular eigenvector must feature in either the first term or the second term of (26c), depending on whether it belongs to $\mathbf{V}_1$ or to $\mathbf{V}_2$. If it belongs to $\mathbf{V}_2$, then potential "wrongness" in the model prediction arises from the fact that the calibration process provides no information that is relevant to that prediction. Potential predictive error is governed entirely by σ_p^2, the inherent (pre-calibration) uncertainty of system properties, for the second term in (26c) is zero because of orthogonality of $\mathbf{y}$ to all members of $\mathbf{V}_1$ (because the eigenvectors comprising the columns of $\mathbf{V}$ are all orthogonal to each other). Thus the uncertainty of this prediction is undiminished from that which prevailed prior to model calibration.

345 On the other hand, if the prediction sensitivity vector **y** is parallel a $\mathbf{V}_1$ eigenvector, the
346 first term of equation 26c is zero and the second term is nonzero. The magnitude of this
347 second term depends on two factors. One is the goodness of model-to-measurement fit as
348 encapsulated in the term σ_h^2; the other is the magnitude of the eigenvalue corresponding
349 to the eigenvector to which the prediction sensitivity vector is parallel. If this eigenvalue
350 is small, the contribution to uncertainty arising from the second term can be very large,
351 possibly larger than if the model had not been calibrated at all. In this case the calibration
352 dataset says less about the parameter combinations that define prediction sensitivity than
353 can be said on the basis of knowledge about system properties from outside of the
354 calibration process altogether. Under these circumstances, once again, the calibration
355 process provides no assistance in reducing predictive uncertainty below that which exists
356 if the model had not been calibrated at all. On the other hand, if the pertinent eigenvalue
357 is large, then the calibration process may reduce predictive error variance substantially,
358 the extent to which it does this being dependent on the measurement error variance σ_h^2.

359 **Tailoring the Calibration Process to Reduce Predictive Error**

360 The notion of a "calibrated model" conveys the idea that a model, once calibrated, can be
361 used to make a variety of different predictions of system behavior. It also suggests that
362 the calibration process is independent of the prediction process. The analysis presented
363 above, however, suggests that the usefulness of environmental models in making critical
364 predictions of system behavior can be enhanced if the calibration process is undertaken
365 with predictions required by the model kept specifically in mind.

When undertaking regularized inversion based on truncated SVD, eigenvectors can be shifted from $\mathbf{V}_2$ to $\mathbf{V}_1$ of equations 26 to increase goodness of fit to a level considered acceptable. This is normally done in order of decreasing respective eigenvalue; that is, eigenvectors corresponding to high eigenvalues (and hence respecting broad scale hydraulic property distributions) are normally shifted to $\mathbf{V}_1$ while those with low eigenvalues (reflecting system detail) are retained in $\mathbf{V}_2$. As eigenvectors are shifted from $\mathbf{V}_2$ to $\mathbf{V}_1$ the first term of equations 26 falls monotonically (due to the fact that it is positive definite), while the second term rises monotonically (for the same reason). The sum of these terms will thus show a minimum, the location of this minimum being dependent on the particular prediction being investigated. However there is no certainty that the number of eigenvalues at which this predictive uncertainty minimum is achieved (ie. the number of eigenvalues, counting from the highest, at which truncation occurs and all further eigenvectors are assigned to $\mathbf{V}_2$ and thus assigned an eigenvalue of zero) corresponds to the number of eigenvalues required to achieve (but not undercut) a suitably defined Φ_n. In fact, as will be demonstrated below, minimization of error variance for a particular prediction may require use of many eigenvalues beyond that which is required to achieve a suitable value for Φ_n and can thus result in what classical analysis would perceive as "overfitting". Furthermore, the truncation level for minimizing the uncertainty of one specific prediction may not be the same as that required to minimize the uncertainty of another prediction, this lending weight to the assertion made above that model calibration should be prediction-specific.

So how can model calibration be "tuned" to the prediction that it must make? A number of options exist. One is to actively seek the minimum in the predictive variance curve,

even if this leads to overfitting according to the classical view of model calibration. Another option is to vary from the traditional practice of ranking eigenvalues in decreasing order of their magnitude when deciding on a level of truncation. This traditional strategy always leads to the loss of low eigenvalues from $\mathbf{V}_2$ before high eigenvalues, regardless of the disposition of corresponding eigenvectors with respect to a model prediction of particular interest. An alternative strategy is to take account of the orientation of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ eigenvectors with respect to predictive sensitivity when deciding on the order of eigenvector excision from $\mathbf{V}_2$. Thus, for example, eigenvector A may have a lower eigenvalue than eigenvector B. However its inner product with $\mathbf{y}$ may be higher. If its eigenvalue is not so low that its presence in the second term of equation 26 results in higher predictive uncertainty than its presence in the first, it should be included in the second term (ie. the $\mathbf{V}_1$ term), thus avoiding excision in the truncation procedure.

An alternative, and simpler, strategy can be employed to increase the likelihood that eigenvectors of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ which are parallel to a key model prediction belong to $\mathbf{V}_1$ rather than to $\mathbf{V}_2$, thus reducing the probability of model error in making that prediction. This strategy is to increase the weights associated with observations whose sensitivities are more aligned with a particular prediction, relative to those that are orthogonal to it. This procedure results in a $\mathbf{Q}$ matrix for which equation 6 no longer applies. The second term in equations 26b and 26c becomes more complicated as a result; equations 32 repeat equations 26 where $C(\boldsymbol{\varepsilon})$ is no longer proportional to $\mathbf{Q}^{-1}$.

$$\sigma^2_{s\text{-}\underline{s}} = \mathbf{y}^t(\mathbf{I} - \mathbf{R})C(\mathbf{p})(\mathbf{I} - \mathbf{R})^t\mathbf{y} + \mathbf{y}^t\mathbf{G}C(\boldsymbol{\varepsilon})\mathbf{G}^t\mathbf{y} \qquad (32a)$$

$$\sigma^2_{s\text{-}\underline{s}} = \mathbf{y}^t\mathbf{V}_2\mathbf{V}^t_2C(\mathbf{p})\,\mathbf{V}_2\mathbf{V}^t_2\mathbf{y} + \sigma_h^2\mathbf{y}^t\mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}_1{}^t\mathbf{X}^t\mathbf{Q}C(\boldsymbol{\varepsilon})\mathbf{Q}\mathbf{X}\mathbf{V}_1\mathbf{E}_1^{-1}\mathbf{V}^t_1\mathbf{y} \qquad (32b)$$

411 $$\sigma^2_{s\text{-}\underline{s}} = \sigma_p^2 \mathbf{y}^t \mathbf{V}_2 \mathbf{V}^t_2 \mathbf{y} + \sigma_h^2 \mathbf{y}^t \mathbf{V}_1 \mathbf{E}_1^{-1} \mathbf{V}^t_1 \, \mathbf{X}^t \mathbf{Q} \mathbf{C}(\boldsymbol{\varepsilon}) \mathbf{Q} \mathbf{X} \mathbf{V}_1 \mathbf{E}_1^{-1} \mathbf{V}^t_1 \mathbf{y} \qquad (32c)$$

412 With $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ thus re-formulated, its eigenvectors and eigenvalues also change. However
413 now the eigenvectors which are more parallel to $\mathbf{y}$ will tend to have higher eigenvalues,
414 and hence will be less likely to be truncated in a calibration procedure that orders
415 eigenvalues in order of decreasing magnitude prior to truncation. Care must be taken in
416 implementing this procedure, however, to ensure that the second term of equations 32 is
417 not unduly amplified by providing high weights to observations that are inherently
418 unreliable.

419 This philosophy of weights assignment violates traditional least squares practice. This
420 traditional practice is based on predictive error variance minimization (Bard, 1974) for
421 overdetermined systems, in which all parameters to which a prediction is sensitive are
422 assumed to be individually estimable through the calibration process. However it is worth
423 noting that the strategy of placing increased emphasis on observation types that most
424 resemble the types of predictions that a model will be required to make, is certainly in
425 harmony with the philosophical underpinnings of manual calibration which is often based
426 on the simple, but effective, premise that "if you can't fit everything, then at least fit the
427 things that matter most". The effectiveness of this strategy will be demonstrated below
428 using a synthetic example.

429 It must also be pointed out that adherence to traditional weights assignment practices that
430 are recommended in works such as Hill (1998) presupposes that $\mathbf{C}(\boldsymbol{\varepsilon})$ is known or can be
431 estimated. Where calibration data noise is dominated by model structural error (as is
432 mostly the case), $\mathbf{C}(\boldsymbol{\varepsilon})$ is not known nor can be easily estimated. Furthermore many

433 contexts such "noise" may show considerable spatial correlation of unknown magnitude.
434 Thus the common practice of assuming independence of measurement errors (and thus a
435 diagonal **Q** matrix) is, in fact, a violation of the precepts espoused in guidelines such as
436 these. In view of this fact, the assignment of weights in a manner that places greater
437 emphasis on observations that are more closely related to key model predictions, is
438 probably no less in violation of these precepts than any other method of weights
439 assignment; however in view of its probably beneficial outcome of reducing predictive
440 error variance, it is probably far more effective.

441 Where regularized inversion is undertaken using methods other than truncated SVD (for
442 example constrained minimization) it is also possible to tailor the model calibration
443 process such that the variance of model predictive error is minimized. In equations 27 the
444 reciprocal of the "regularization weight factor" β plays a similar role to that of k, the
445 eigenvalue truncation number employed by the truncated SVD method; higher values of
446 β result in smoother calibrated fields, and higher values of Φ. Normally a value of β is
447 sought which results in an "adequate" level of model-to-measurement misfit, that is, an
448 objective function equal to a suitably chosen Φ_n. However, as will be demonstrated
449 below, upward variation of β results in a monotonic lowering of the second term of
450 equation 26a and a monotonic rise in the first term. Hence there is a value of β for which
451 the error variance of a particular prediction is minimized. This value will rarely coincide
452 with that required to exactly achieve Φ_n. Hence, one option for tuning the calibration
453 process for prediction optimization is to seek that β which minimizes predictive error
454 variance. Alternatively (or as well), higher weights could be assigned to those members
455 of a calibration dataset that are most pertinent to a particular prediction, as was discussed

above in relation to truncated SVD. A further strategy may be to tailor the regularization constraint matrix **Z** to best accommodate the prediction that the model is required to make.

The Effect of Parameter Lumping

Classical approaches to model calibration undertake "pre-emptive" or "implicit" regularization using some form of parameter "lumping". For a groundwater model this often takes the form of spatial parameter definition using a limited number of zones of piecewise parameter uniformity. This approach to regularization has the advantage that, if zones are few enough and defined in accordance with spatial data density, numerical stability of the inversion process is guaranteed. However it has the disadvantage that some of the information content of the calibration dataset may be lost because the chosen parameterization methodology has no means of expressing hydraulic property complexity that may become apparent through the calibration process itself. (It should be noted in passing that regularized inversion and geologically-based parameter zonation are not mutually exclusive; see, for example, De Groot-Hedlin and Constable (1990). The advantage of combining zones with regularized inversion based on a large number of parameters is that intra-zonal heterogeneity can be accommodated in the model at the same time as inter-zonal hydraulic property contrasts if the calibration dataset provides a strong enough indication that such intra-zonal complexity exists.)

Suppose that instead of estimating m parameters whose true values are encapsulated in the vector **p**, j "lumped" parameters comprising the elements of a smaller vector $\underline{\mathbf{r}}$ are

estimated in their stead. Suppose further that model outputs corresponding to observations are calculated using the relationship:-

$$\mathbf{o} = \mathbf{W}\underline{\mathbf{r}} \tag{33}$$

$\underline{\mathbf{r}}$ can be determined through objective function minimization using the formula (same as equation 8):-

$$\underline{\mathbf{r}} = (\mathbf{W}^t\mathbf{Q}\mathbf{W})^{-1}\mathbf{W}^t\mathbf{Q}\mathbf{h} \tag{34}$$

Substitution of (4) then yields:-

$$\underline{\mathbf{r}} = (\mathbf{W}^t\mathbf{Q}\mathbf{W})^{-1}\mathbf{W}^t\mathbf{Q}(\mathbf{X}\mathbf{p} + \boldsymbol{\varepsilon}) \tag{35a}$$

$$= \mathbf{R}'\mathbf{p} + \mathbf{G}'\boldsymbol{\varepsilon} \tag{35b}$$

where the "modified resolution matrix" $\mathbf{R}'$ and the matrix $\mathbf{G}'$ of equations 35 are given by:-

$$\mathbf{R}' = (\mathbf{W}^t\mathbf{Q}\mathbf{W})^{-1}\mathbf{W}^t\mathbf{Q}\mathbf{X} \tag{36a}$$

and

$$\mathbf{G}' = (\mathbf{W}^t\mathbf{Q}\mathbf{W})^{-1}\mathbf{W}^t\mathbf{Q} \tag{36b}$$

Let the relationship between a parameterization based on many parameters $\underline{\mathbf{p}}$, and that based on lumping of these parameters (i.e. $\underline{\mathbf{r}}$) be described by the equation:-

$$\underline{\mathbf{p}} = \mathbf{L}\underline{\mathbf{r}} \tag{37}$$

where **L** is a $m \times j$ "lumping matrix". In many cases each row of **L** will be comprised of zero elements except for a single element of 1. For example if zone-specific pilot points (see Doherty, 2003) are employed as the spatial parameterization basis for $\underline{\mathbf{p}}$, while the elements of $\underline{\mathbf{r}}$ are zonal parameter values, each element of $\underline{\mathbf{p}}$ will have the same value as the element of $\underline{\mathbf{r}}$ that pertains to the zone in which the corresponding pilot point lies.

With **L**, **R**' and **G**' defined as above, it is easy to show that **R** and **G** matrices for use in equation 26a can be calculated as:-

$$\mathbf{R} = \mathbf{L}\mathbf{R}' \tag{38a}$$

and

$$\mathbf{G} = \mathbf{L}\mathbf{G}' \tag{38b}$$

With these definitions of **R** and **G**, equation 26a provides the full expression for predictive error variance of a lumped parameter model. For reasons already discussed in relation to underdetermined systems, omission of the first term of this equation (as is usually done in practice) can lead to significant underestimation of predictive error variance, especially where the underlying system is complex. Unfortunately, however, equation 26a may be difficult to apply in practice. This is because its evaluation assumes that even though only lumped parameters are estimated, sensitivities of a much larger number of distributed parameters have also been calculated (for use in the **X** matrix). Nevertheless, this analysis demonstrates that estimation of predictive error variance based only on the statistics of model-to-measurement misfit neglects an extremely important contributor to potential model error. Where lumping is significant, and **R** is thus

515 significantly different from **I**, and/or where predictive sensitivities have a large
516 component in the subspace of m dimensional parameter space spanned by (**I** – **R**), neglect
517 of this contribution may make such estimates almost meaningless.

518 ## Synthetic Example

519 ### Model Description

520 The principals discussed in the preceding section are illustrated with reference to a
521 synthetic model. This is the same model as that used by Moore and Doherty (2004a and
522 2004b) in an earlier discussion of the use of regularized inversion in groundwater model
523 calibration.

524 Figure 1a shows the 500m × 800m rectangular domain of a single layer groundwater
525 model of flow in a confined aquifer of 10m thickness. A fixed inflow of $0.1/m^3/day/m$
526 occurs through the upper boundary of the model; heads are fixed at 0m along the lower
527 boundary. An hydraulic conductivity field with a log average value of zero was generated
528 using a log-exponential variogram with a range (3 times the coefficient in the exponent in
529 the variogram equation) of 200m and a sill of 0.2. Diffuse recharge is zero. Flow within
530 the domain was simulated using MODFLOW-2000 (Harbaugh et al, 2000) using a finite-
531 difference grid consisting of 50 rows and 80 columns of 10m square cells. The travel time
532 and track of a particle released near the top boundary was simulated using the ADV
533 package of MODFLOW-2000 (Anderman and Hill, 2001); the path of the particle is
534 depicted in Figure 1a.

535 On the basis of the hydraulic conductivity field shown in Figure 1a, heads were generated
536 at 12 wells, the locations of which are shown in Figure 1b; heads in these wells vary
537 between 5.7m in the upper part of the model domain and 1.1m in the lower part of the
538 domain. These heads were used for model calibration after the addition of Gaussian noise
539 with a standard deviation of 0.3m.

540 Spatial parameterization was implemented using pilot points. As described in Doherty
541 (2003) and Moore and Doherty (2004a), the calibration process assigns hydraulic
542 conductivity values to these points; these values are then spatially interpolated to all cells
543 of the model domain (using kriging in the present case).

544 Regularized inversion of the head data was undertaken using PEST (Doherty, 2004), with
545 assistance from the PEST Groundwater Utilities (Doherty, 2003); both truncated SVD
546 and constrained minimization regularization were employed. In the latter case
547 regularization constraints were of the "preferred value" type, with the log of each pilot
548 point hydraulic conductivity being assigned a preferred value of zero; thus the $\mathbf{Z}$ matrix
549 of equation 27 was an $m \times m$ identity matrix. $\mathbf{Q}_t$ was calculated as the inverse of an inter-
550 pilot point covariance matrix; pilot point covariances were calculated using the same
551 variogram as that employed for generation of the hydraulic conductivity field. The
552 regularization weight factor β was calculated by PEST as that required to achieve a user-
553 supplied value for Φ_n, the objective function at which "adequate calibration" is deemed to
554 occur in accordance with the level of measurement noise.

555 Analyses in this section focus on a prediction of the particle exit location, the "true value"
556 of this prediction being 206.78 meters from the left side of the model. This prediction,

557 rather than travel time, was chosen for the analyses described below due to the fact that
558 the latter prediction is relatively unconstrained by a calibration process that is based on
559 heads alone.

560 **Predictive Variance Minimization**

561 With Gaussian noise of standard deviation 0.3m added to the heads, an objective function
562 value (i.e. Φ_n of equation 9) of 12.0 should be sought in a regularized inversion process in
563 which measurement weights are all set to 3.33, this being the inverse of the noise
564 standard deviation. (Note that weights are squared to form the diagonal elements of **Q**.)

565 Table 1 summarizes the outcomes of undertaking regularized inversion using truncated
566 SVD with a varying number of pre-truncation eigenvalues. It is apparent from this table
567 that an objective function of 12 can be achieved with as few as 4 eigenvalues. If the terms
568 of equations 26 are computed using sensitivities calculated on the basis of "calibrated"
569 parameters (i.e. parameters calculated using 4 eigenvalues), the graphs of Figure 2 result;
570 note that C(**p**) in equations 26 was calculated using the same variogram as that which was
571 used to generate the hydraulic conductivity field, and thus properly represents the spatial
572 characteristics of the true hydraulic property field. Figure 3a shows the "calibrated"
573 parameter field.

574 The monotonic rise of the second term of equations 26, and the monotonic fall of the first
575 term are apparent from Figure 2 (except for the first point, where numerical noise
576 incurred through the SVD process causes a slight rise). As the number of eigenvalues
577 increases, the second term does not rise fast enough relative to the first term for the sum
578 of the two terms to incur a minimum; however if the graph were to be extended to 13

579 eigenvalues (at which stage the second term of equation 26 would be extremely high due
580 to the fact that with 12 observations the rank of $\mathbf{X}^t\mathbf{Q}\mathbf{X}$ is only 12), a pronounced
581 minimum at 12 eigenvalues would be apparent.

582 Figure 2 demonstrates that even though the model can be considered to be calibrated with
583 truncation occurring at 4 eigenvalues, minimum predictive error variance is achieved at
584 12 eigenvalues. It is interesting to note from Table 1 that the most accurate prediction of
585 particle exit point location was made by the model calibrated using 7 eigenvalues.

586 The calibration process was repeated using constrained minimization regularization. For
587 Φ_n equal to 12, the corresponding β value is 1.76. The resulting calibrated parameter field
588 is shown in Figure 3b.

589 Using sensitivities calculated on the basis of the calibrated parameter field, the terms of
590 equation 26 were calculated for varying values of β; see Figure 4. Monotonicity of the
591 two terms of equation 26 is clearly demonstrated in this figure. It is also apparent that the
592 minimum predictive error variance is achieved at a β value of about 1.0. This
593 demonstrates once again that model performance with respect to the prediction of particle
594 exit location is optimized when the calibration process involves some degree of
595 overfitting.

596 Table 2 shows model-predicted exit points calculated on the basis of parameters achieved
597 through regularized inversion using different values of β. The closest prediction to the
598 actual exit point occurs at a β value of between 0.86 and 1.12.

Observation Weights Adjustment for Prediction Optimization

In the previous section it was demonstrated that a model can be considered "calibrated", yet may not be optimally parameterized for the making of a specific prediction. It was suggested in the theoretical section of this paper that if data is weighted in accordance with its relevance to a specific prediction, the calibrated model may be capable of making that prediction with a lower probability of error.

For the analyses documented in the previous section, observation weights were uniformly set at 3.33, this being the inverse of measurement uncertainty. These analyses were repeated with weights for 5 of the 12 observations being doubled. With this weighting strategy the model can be considered "calibrated" with an objective function of 27.0. The observations chosen for enhanced weighting were head values at the wells depicted as diamonds in Figure 1b. These observations were chosen for special treatment because the normalized inner products of the prediction sensitivity with the sensitivities of these observations were all greater than 0.1; for all other observations normalized inner products were less than 0.1. (Inner products were calculated using sensitivities pertaining to the parameters depicted in Figure 3b.)

Table 3 shows the results of truncated SVD regularized inversion with a varying truncation limit. Once again, the model can be said to be "calibrated" if only 4 eigenvalues are employed. Figure 5 shows the terms of equation 32 (which must now be used instead of (26) because equation 6 no longer applies) calculated using sensitivities pertaining to the "calibrated model". The minimum predictive error variance (once again obtained at 12 eigenvalues) is slightly lower than that obtained with uniform weights.

621 What is more important however, is that the predictive error variance at 4 eigenvalues is
622 much lower than for uniform weights. Thus the "calibrated model" is a much better
623 predictor of the exit point location. Calibration results listed in Table 3 support this
624 conclusion.

625 The worth of selective observation weights enhancement was also tested using
626 constrained minimization regularization. Figure 6 and Table 4 show the results.
627 Calibration is achieved at a β value of about 3.3 (i.e. with $1/\beta$ equal to about 0.3).
628 Predictive error variance at the point of calibration is lower in Figure 6 than in Figure 4
629 where no weights enhancement was undertaken. Table 4 shows that the "calibrated
630 model" is indeed a good predictor of particle exit point location.

631 **Predictive variance with parameter zonation**

632 As described above, predictive error variance calculation based on lumped
633 parameterization can be accommodated if equation 26a incorporates **R** and **G** matrices
634 described by equations 38. This is now demonstrated using parameters based on zones of
635 piecewise uniformity.

636 Figure 7 shows the zonation pattern chosen for this demonstration; boundaries for three
637 (non-contiguous) zones were drawn against the background of the true hydraulic
638 conductivity field depicted in Figure 1a (a luxury not available in normal modeling
639 practice). Estimation of zonal hydraulic conductivities was undertaken using the same
640 calibration dataset as that used for previous analyses. However weights enhancement was
641 not employed, each head measurement thus being assigned a weight of 3.33. The **X**
642 matrix of equation 26 was calculated on the basis of the pilot points depicted in Figure

643 1b, with "calibrated" values assigned to these points according to the zone in which they
644 lie; however, in order to establish true correspondence between pilot point and zone
645 parameters, these sensitivities were calculated on the basis of a kriging procedure which
646 prevented interpolation from pilot points in one zone to cells occupying another zone.

647 Predictive error variance was calculated as $10189m^2$. The contribution to this variance
648 from the second term of equation 26 is $173m^2$, this being the predictive error variance
649 that would have been calculated using traditional methods that ignore the implicit
650 regularization of the lumping process. The actual model prediction was a surprisingly
651 good 210m. This level of accuracy is partly attributable to luck, and partly attributable to
652 the fact that zonation definition took place against the background of the true hydraulic
653 conductivity field, this constituting the addition of a significant amount of extra
654 information to the calibration process. Nevertheless, the above calculation shows that the
655 potential error associated with this model prediction is still quite high.

656 **Discussion and Conclusions**

657 It has been demonstrated that where a model must be used to make a specific prediction,
658 or type of prediction, it can become a better instrument for the making of that prediction
659 if this role is borne firmly in mind during the process of calibrating it. One way in which
660 this can be achieved is to give more weight to observations that most resemble the type of
661 predictions the model is required to make. Though demonstrated in the context of a small
662 synthetic groundwater model, this approach to calibration can be extended to other model
663 types as well. For example if a watershed model is to be employed for prediction of flood
664 peaks, then it will probably be a better predictor of those peaks if a weighting philosophy

665 is chosen that allows the rising limbs of historical flow peaks, and the peaks themselves,
666 to be "highly visible" when the model is calibrated. Alternatively, if the model is to be
667 used for the making of predictions pertinent to low flow conditions (such conditions often
668 being important for biotic health), then historical recessions and low flows should figure
669 prominently in the calibration process.

670 This philosophy becomes especially important where a model is used to predict the
671 response of an environmental system to extreme events (which many models are built
672 specifically to do). Under these circumstances, the historical data that most resembles the
673 types of predictions which the model will be required to make may comprise only a small
674 part of the total calibration dataset. If this data is not given sufficient weight to be clearly
675 "seen" during the calibration process, in spite of the numerical predominance of other
676 data, then the model will not live up to its full potential as an adequate predictor of these
677 extremes. This pragmatic approach to weights assignment is significantly different from
678 that which seeks to make weights reflective of measurement (or inferred structural) noise,
679 and/or that which seeks to promote homoscedascity of residuals through appropriate
680 measurement and model output transformation, the latter being common practice in
681 stream flow modeling (see, for example, Box and Cox, 1964 and the many papers which
682 reference this work). Nevertheless, because of its intuitive appeal, it has been a firm tenet
683 of manual model calibration for many years. Furthermore, flexibility and pragmatism in
684 weights assignment, especially as this pertains to the assignment of relative weighting
685 between different measurement types, has received some positive treatment in the
686 modeling literature; see for example Weiss and Smith (1998) and Gupta et al. (1998).

687 The suggested prediction-specific nature of the calibration process has ramifications for
688 the way models are viewed, both by modelers themselves, and by members of the broader
689 community that rely on them. Models are primarily viewed as numerical "emulators" of
690 environmental processes. A particular model (especially a complex model), may have
691 many different types of output, each pertaining to different system state variables.
692 Anyone who has ever built a model knows that it would be very fortunate indeed if all
693 model outputs were to replicate all corresponding system states all the time, even after
694 large amounts of effort have been devoted to calibrating it. This simply does not happen
695 because numerical models are always inadequate simulators of system behavior, with
696 process equations, and the parameters that they use, being simplifications of reality.
697 Because of these simplifications, parameters often take on "hidden roles", assuming
698 values which are not always reflective of the system properties that they purport to
699 represent, but which nevertheless result in more acceptable fits between model outputs
700 and field measurements. This is often vividly demonstrated when a resolution matrix **R**
701 (any row of which describes the contribution of different real-world parameter values to a
702 single calibrated model parameter value) displays the phenomenon of "parameter
703 contamination"; this occurs when part of the value of a calibrated parameter is
704 contributed by real-world parameters of entirely different type.

705 So it cannot be denied that because it is a simplified simulator of real-world processes,
706 that employs parameters that are incapable of representing the intrinsic heterogeneity of
707 natural systems, and that are estimated through calibration against inadequate and noisy
708 datasets, a model will yield erroneous predictions of future environmental behavior at
709 some places for some of the time. Nor can it be denied that even the most complex

710 physically-based model is not immune from some of the properties of a "black box", as
711 parameters are adjusted to ensure that the extent of model error is minimized (or at least
712 made "sufficiently good") under historical conditions. This should not detract from a
713 properly-constructed model's status as the most suitable means to process a given
714 environmental dataset. However perhaps it does mean that a model should probably be
715 regarded less as a potentially perfect simulator of environmental behavior and more as a
716 sophisticated processor of environmental data, especially when used in conjunction with
717 high-end inversion software. Viewed in that capacity, this role is best served when
718 maximum information content is extracted from a given environmental dataset. Its
719 limitations as an imperfect simulator of reality dictate, however, that it may not be
720 possible to extract the entirety of this information in any one calibration exercise.
721 However, if the calibration process is properly designed in relation to a specific
722 prediction which the model will be required to make, then it can be "tuned" to extract the
723 entirety of the information content of a calibration dataset as it pertains to that specific
724 prediction, thereby endowing it with an enhanced ability to make that prediction
725 accurately under future conditions.

726 **References**

727 Anderman, E.R. and Hill, M.C., MODFLOW-2000, the U.S. Geological Survey modular
728 ground-water model -- Documentation of the ADVective-Transport observations (ADV2)
729 Package: U.S. Geological Survey Open-File Report 01-54, 69p. 2001.

730 Backus, G. and Gilbert, F., The resolving power of gross earth data, *Geophys. J. Roy.*
731 *Astr. S* .,16,169-205, 1969.

732 Bakr, M.I., te Stroet, C.B.M. and Meijerink, A., Stochastic groundwater quality
733 measurement: Role of spatial variability and conditioning, *Water Resour. Res.*, 39(4),
734 1078, doi:10.1029/2002WR001745, 2003.

735 Bard, Y., Nonlinear parameter estimation. New York: Academic Press; 1974.

736 Box, G.E.P. and Cox., D.R., An analysis of transformations, *J.R. Stat. Soc. Ser. B*, 26,
737 211-243, 1964.

738 Christensen, S. and Cooley, R.L., Evaluation of prediction intervals for expressing
739 uncertainties in groundwater flow predictions, *Water Resour. Res*, 35(9), 2727-2639,
740 1999

741 Clemo, T., Michaels, P. and Lehman, R.M., Transmissivity Resolution Obtained from
742 the Inversion of Transient and Pseudo-Steady Drawdown Measurements. MODFLOW
743 and More 2003: Understanding through Modeling - Conference proceedings 2003;
744 Poeter, Zheng, Hill & Doherty. Downloadable from www.mines.edu/igwmc. 2003.

745 De Groot-Hedlin, C. and Constable, S., Occam inversion to generate smooth, 2-
746 dimensional models from magnetelluric data, *Geophysics* , 55(12),1613-1624, 1990.

747 Doherty, J., Groundwater model calibration using pilot points and regularization, *Ground*
748 *Water*, 41 (2), 170-177, 2003.

749 Doherty, J., PEST: Model-Independent Parameter Estimation. Version 5 of User manual.
750 Watermark Numerical Computing, Brisbane, Australia, 2004.

751 El Harrouni, K., Ouazar, D., Wrobel, L.C. and Cheng, A.H-D., Uncertainty Analysis in
752 groundwater flow with DRBEM, *Engineering Analysis with Boundary Elements* 19 , 217-
753 221, 1997.

754 Feyen, L., Gómez-Hernández, J.J., Ribeiro, P.J., Beven. K.J., De Smedt. F., A
755 Bayesian approach to stochastic capture zone delineation incorporating tracer arrival
756 times, conductivity measurements, and hydraulic head observations, *Water Resour. Res.*
757 39(5), sbh (10)1-13, 2003.

758 Franssen, H.H., Gómez-Hernández, J.J., Sahuquillo, A., Coupled inverse modelling of
759 groundwater flow and mass transport and the worth of concentration data, *J. Hydrol.*,
760 281:281-295, 2003.

761 Frederick, D. and Doherty, J., Regularized inversion using head and concentration
762 measurements. MODFLOW and More 2003: Understanding through Modeling -
763 Conference proceedings 2003; Poeter, Zheng, Hill & Doherty. Downloadable from
764 www.mines.edu/igwmc, 2003.

765 Gómez-Hernández, J.J., Hendricks Fransen, H.J., and Sahuquillo, A., Stochastic
766 conditional inverse modeling of subsurface mass transport: A brief review of the self-
767 calibrating method, *Stoch. Env. Res. Risk. A.*, 17:319-328, 2003.

768 Gorokhovski, V,M., Problem-Dependence of Ground-Water Model Identification:
769 Significance, Extent, and Treatment, *Ground Water*, 34(3):461-469, 1996.

770

771

772 Guadagnini A and Neuman SP. Nonlocal and localized analyses of conditional mean
773 steady state flow in bounded, randomly uniform domains, 1, theory and computational
774 approach, *Water Resourc. Res.* , 35(10):2999-3018, 1999.

775 Gupta, H.V., Sorooshian, S. and Yapo, P.O., Towards improved calibration of hydrologic
776 models: multiple and non-commensurable measures of information. *Water Resour. Res.*,
777 34 (4), 751-763, 1998.

778 Harbaugh, A.W., Banta, E.R., Hill, M.C. and McDonald. M.G., The U.S. Geological
779 Survey Modular Ground-Water Model – User Guide to Modularization Concepts and the
780 Ground-Water Flow process. U.S. Geological Survey Open File Report 00-92. Reston,
781 Virginia, 2000.

782 Harvey, C.H. and Gorelick, S.M.., Mapping Hydraulic Conductivity – Sequential
783 Conditioning with Measurements of Solute Arrival Time, Hydraulic Head, and Local
784 Hydraulic Conductivity, *Water Resourc. Res.*, 31(7),1615-1626, 1995.

785 Hendricks-Franssen, H.J., Gomez-Hernandez, J. and Sahuquillo, A., Coupled inverse
786 modeling of groundwater flow and mass transport and the worth of concentration data, *J.*
787 *Hydrol.*, 281(4),281-295, 2003.

788 Hill, M.C., Analysis of accuracy of approximate, simultaneous, nonlinear confidence
789 intervals on hydraulic heads in analytical and numerical cases, *Water Resour. Res.*;
790 25(2),177-190, 1989.

791 Hill, M.C., Methods and guidelines for effective model calibration. U.S. Geological
792 Survey, Water-Resources Investigations Report 98-4005. 1998.

793 Kitanidis, P.K., The minimum structure solution to the inverse problem, *Water Resour.*
794 *Res.* 33(10),2263-2272, 1997.

795 LaVenue, A.M., RamaRao, B.S., de Marsily, G. and Marietta, M.G., Pilot point
796 methodology for automated calibration of an ensemble of conditionally simulated
797 hydraulic conductivity fields 2, Application, *Water Resour. Res.* 1995;31(3):495-516.

798 Menke, W., Geophysical Data Analysis: Discrete Inverse Theory. Academic Press Inc;
799 1984.

800 Moore, C. and Doherty, J., Regularized Inversion in Groundwater Model Calibration.
801 Part 1. The Cost of Uniqueness. Submitted for Publication in Advances in Water
802 Resources, 2004a.

803 Moore, C. and Doherty, J., Regularized Inversion in Groundwater Model Calibration.
804 Part 2. True and Estimated Heterogeneity. Submitted for Publication in Advances in
805 Water Resources, 2004b.

806 Poeter, E.P. and Belcher, W.R., Assessment of Porous Media Heterogeneity by Inverse
807 Plume Analysis, *Ground Water*, 29(1),56-62, 1991.

808 RamaRao, B.S., LaVenue, A.M., de Marsily, G. and Marietta, MG., Pilot point
809 methodology for automated calibration of an ensemble of conditionally simulated
810 hydraulic conductivity fields, 1, theory and computational experiments, *Water Resour.*
811 *Res.*, 31(3),475-493, 1995.

812 Tiedeman, C.R., Hill, M.C., D'Agnese, F.A., and Faunt, C.C., Methods for using
813 groundwater model predictions to guide hydrogeologic data collection, with application
814 to the Death Valley regional ground-water flow system: *Water Resour. Res.*,
815 39(1):10.1029/2001 WR001255, 2003.

816 Vasco, D.W., Datta-Gupta, A. and Long, J.C.S., Resolution and uncertainty in hydrologic
817 characterization. *Water Resour. Res.*;33(3):379-397,1997.

818 Vecchia, A.V. and Cooley, R.L., Simultaneous confidence and prediction intervals for
819 nonlinear regression models with application to a groundwater flow model. *Water*
820 *Resour. Res.* 23 (7), 1237-1250, 1987.

821 Weiss, R. and Smith, L., Parameter space methods in joint parameter estimation for
822 groundwater flow models, *Water Resour. Res.*, 34(4), 647-661, 1998.

823 Wiggins, R.A., Larner, K.L. and Wisecup, R., Residual statics analysis as a general linear
824 inverse problem, *Geophysics*, 41(5),922-938,1976.

825 Woodbury, A.D. and Ulrich, T.J., A full-bayesian approach to the groundwater inverse
826 problem for steady state flow, *Water Resour. Res.*,36(8),2081-2093, 2000.

827 Woodbury, A.D. and Rubin,Y., A full-bayesian approach to parameter inference from
828 tracer travel time moments and investigation of scale effects at the Cape Cod
829 experimental site, *Water Resour. Res.*, 36(1):159-171, 2000.

List of Tables and Figures

Tables

Figures

Table 1. Results of truncated SVD inversion with varying number of eigenvalues.

Eigenvalues before truncation	Objective Function (m^2)	Exit point prediction (m)
1	34.71	245.3
2	21.09	244.7
3	13.08	257.9
4	11.18	251.9
5	11.19	247.3
6	5.5	264.2
7	3.15	187.0
8	3.087	172.3
9	3.35	183.5
10	1.64	123.4
11	3.155E-3	159.5
12	1.8E-11	155.7

Table 2. Results of constrained minimization inversion using different regularization weight factors.

β	Objective Function (m^2)	Exit point prediction (m)
11.84	30.0	220.5
5.66	24.0	242.1
2.372	15.0	233.1
1.76	12.0	226.9
1.41	10.0	244.3
1.12	8.0	211.1
0.86	6.0	197.0
0.5317	3.0	163.9
0.3046	1.0	140.6
0.13	0.1	154.1

Table 3. Results of truncated SVD inversion with varying number of eigenvalues; measurement weights doubled for wells shown as diamonds in Figure 1b.

Eigenvalues before truncation	Objective Function (m^2)	Exit point prediction (m)
1	73.74	244.9
2	37.35	240.2
3	29.69	245.2
4	20.12	220.3
5	13.95	205.4
6	13.57	204.9
7	6.847	203.9
8	3.933	146.6
9	3.919	158.2
10	1.363	138.9
11	0.399	143.4
12	2.33e-11	155.7

Table 4. Results of constrained minimization inversion using different regularization weight factors; measurement weights doubled for wells shown as diamonds in Figure 1b.

β	Objective Function (m^2)	Exit point prediction (m)
6.482	42	230.9
4.162	32	218.8
3.305	27	210.7
2.588	22	201.3
1.412	12	177.1
0.8309	6	151.3
0.5536	3	131.9
0.3225	1	124.2

Figure 1. (a) Hydraulic conductivity distribution within rectangular model domain and trace of released particle. (b) Locations of observation wells (circles and diamonds) and pilot points (crosses).

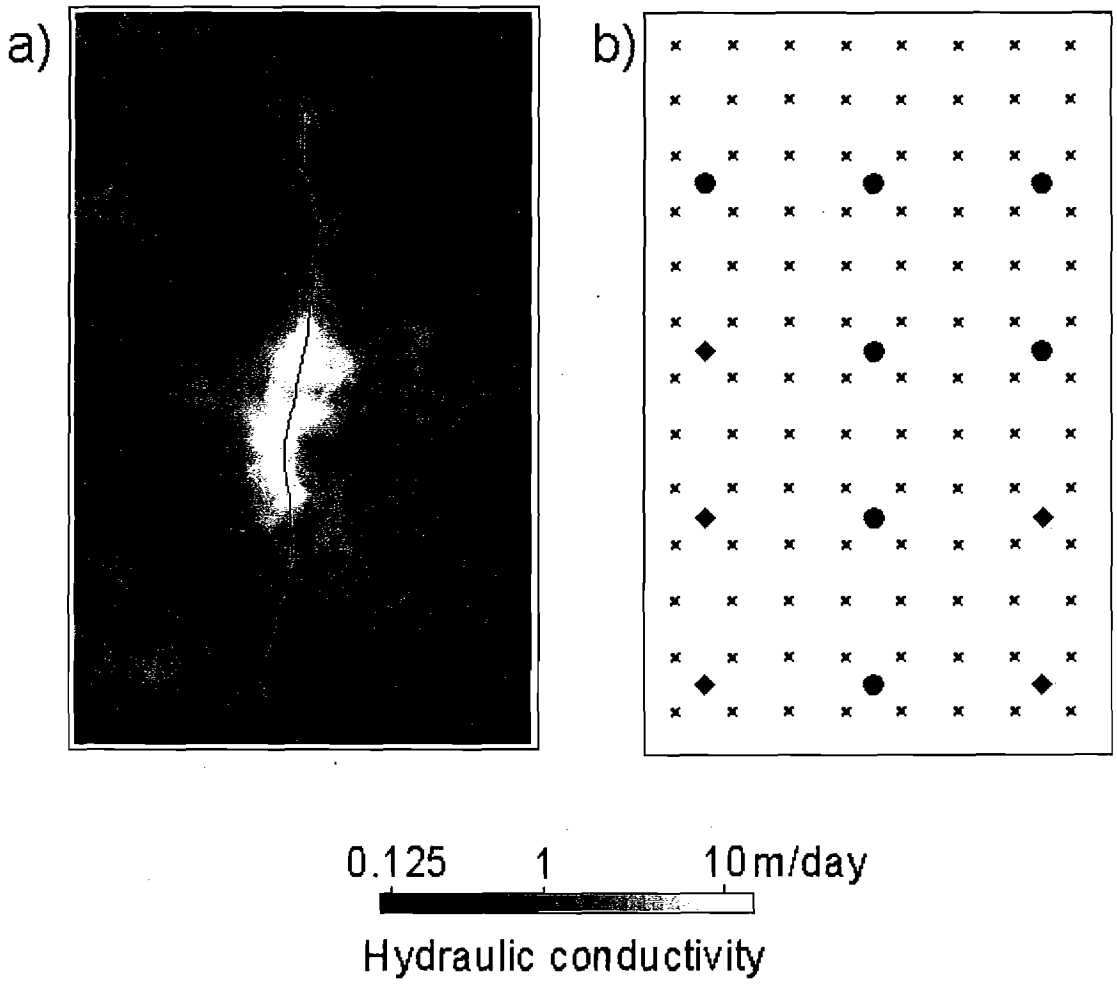


Figure 2. Terms of equation 26 and total model predictive error variance vs. number of eigenvalues. Sensitivities were calculated on the basis of parameters estimated using 4 eigenvalues.

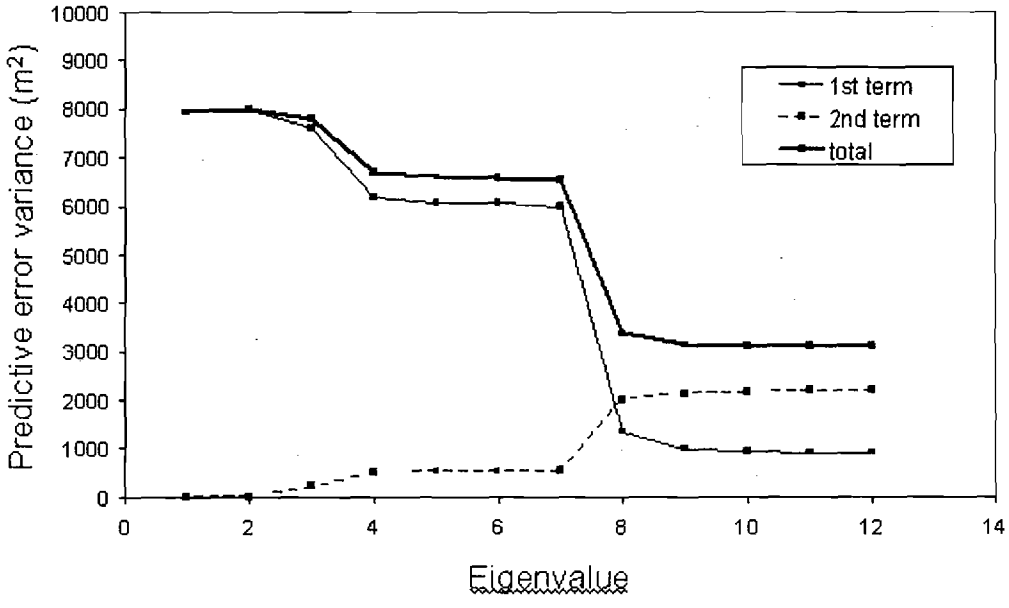


Figure 3. Calibrated hydraulic conductivity fields calculated using (a) truncated SVD with 4 eigenvalues and (b) constrained minimization regularization with Φ_n set to 12.0.

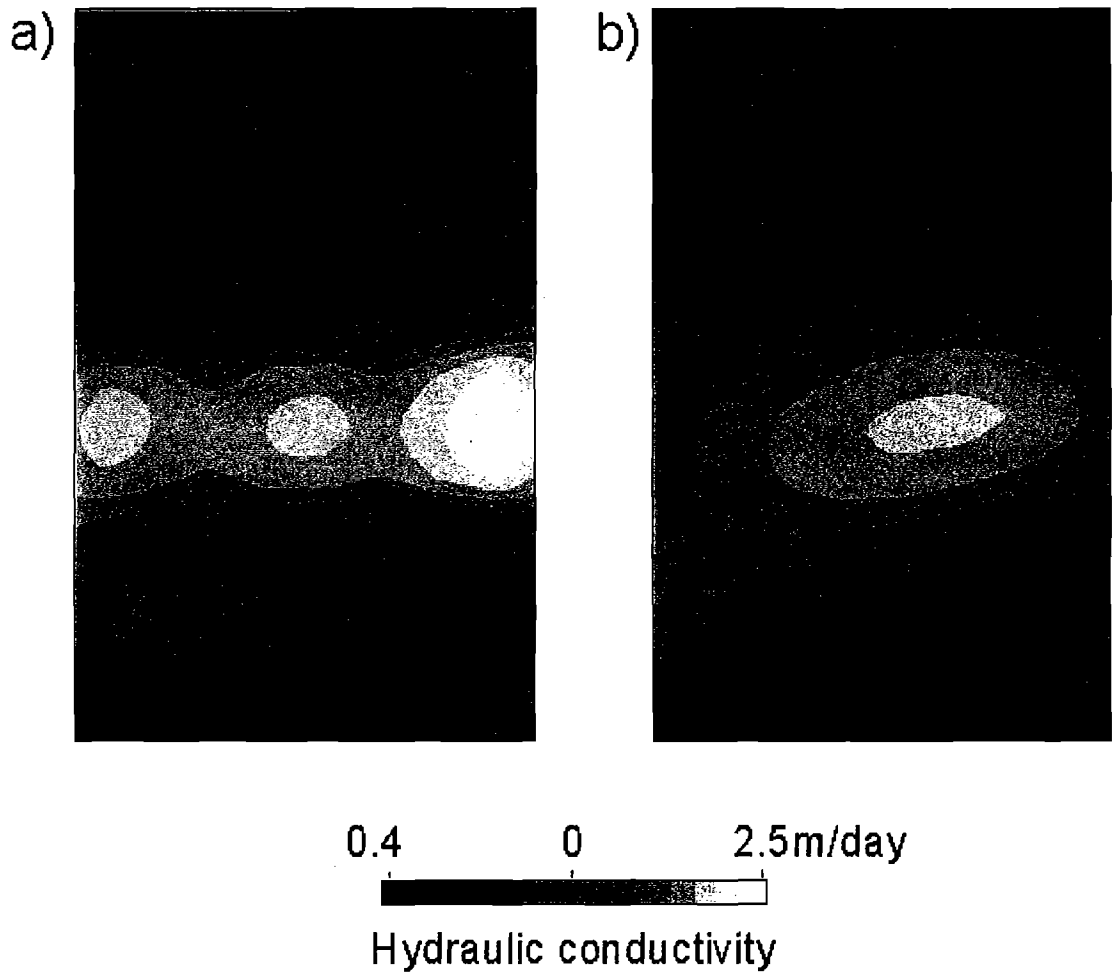


Figure 4. Terms of equation 26 and total model predictive error variance vs. 1/β. Sensitivities were calculated on the basis of parameters estimated using a β value of 1.76 (i.e. a 1/ β value of 0.57).

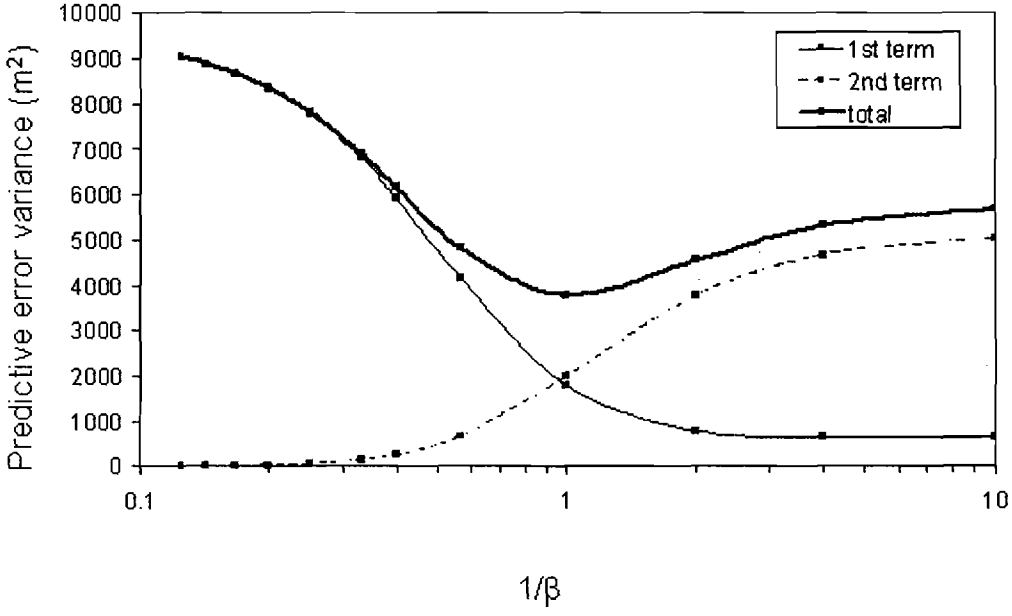


Figure 5. Terms of equation 32 and total model predictive error variance vs. number of eigenvalues. Sensitivities were calculated on basis of parameters estimated using 4 eigenvalues. Weights doubled at wells shown as diamonds in Figure 1b.

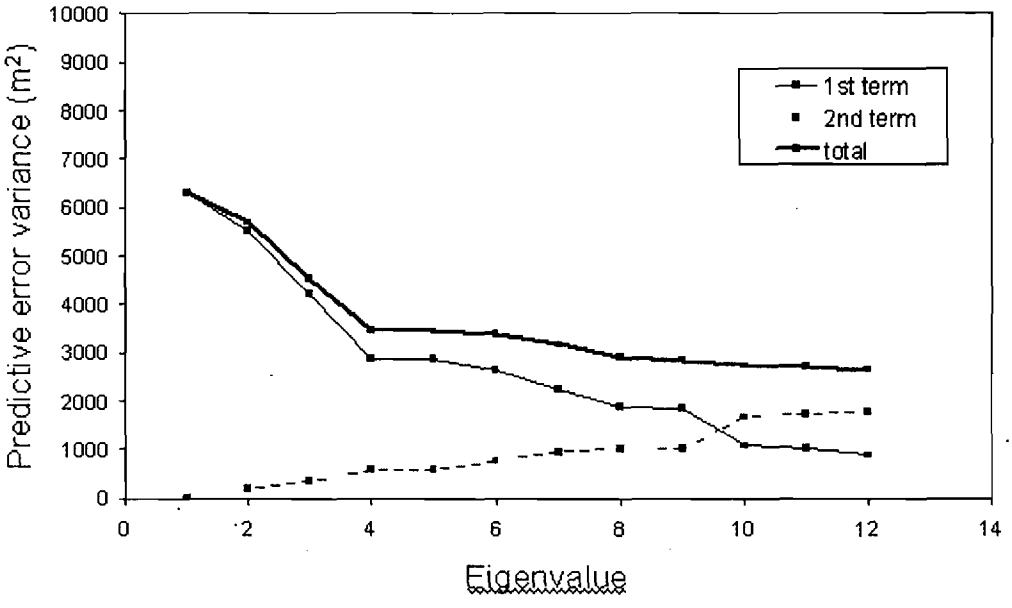


Figure 6. Terms of equation 26 and total model predictive error variance vs. 1/β. Sensitivities were calculated on the basis of parameters estimated using a β value of 3.3 (i.e. a 1/ β value of 0.3).

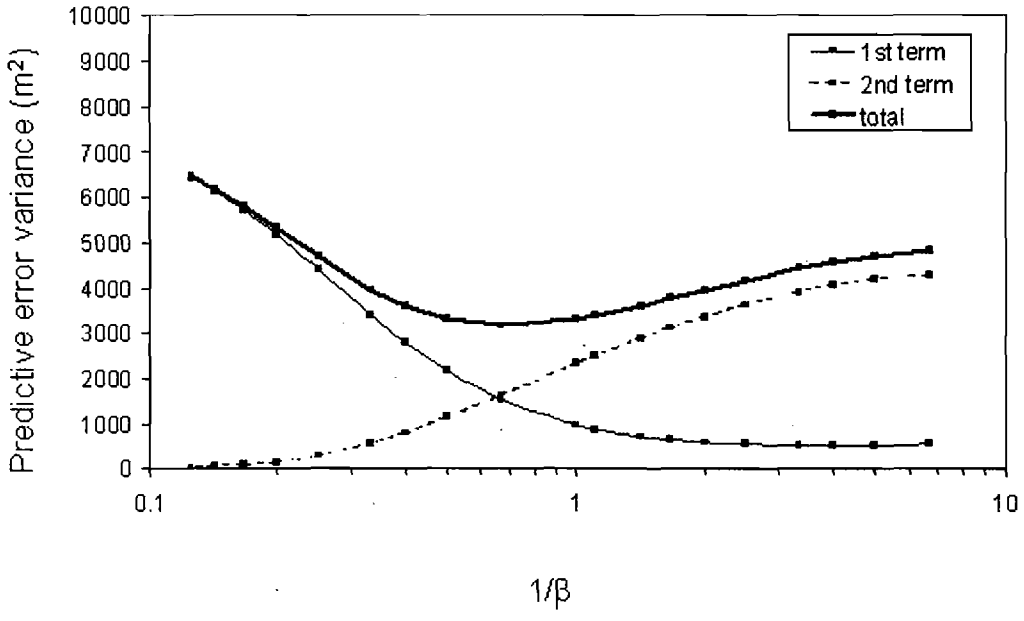


Figure 7. Model domain showing hydraulic conductivity field determined using zones; predicted particle track line is also shown.